GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: January 31, 2003, 03:49:05 ; Search time 308 Seconds
     (without alignments)
     551.370 Million cell updates/sec

Title:
Perfect score: 378
Sequence: 1 tacggctgcttaagacgact............taccattggagaaccatgcg 378

Scoring table: OLIGO_NUC
     GapOp 60.0 , Gapext 60.0
Searched: 396772 seqs, 224632407 residues
Word size : 0

Total number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	19	5.0	527	10	US-09-864-761-7775	Sequence 7775, Ap
2	19	5.0	709	10	US-09-910-943-287	Sequence 287, App
3	18	4.8	26555	9	US-09-860-670-161	Sequence 161, App
4	17	4.5	309	9	US-09-796-692-35	Sequence 35, Appl
5	17	4.5	309	10	US-09-796-692-3533	Sequence 3533, Ap
6	17	4.5	397	10	US-09-878-574-3537	Sequence 3537, Ap
7	17	4.5	627	9	US-10-040-739-843	Sequence 843, App
8	17	4.5	837	9	US-09-938-842A-1704	Sequence 1704, Ap
9	17	4.5	1005	10	US-09-867-550-1051	Sequence 1051, Ap
10	17	4.5	1041	10	US-09-864-761-31616	Sequence 31616, A
11	17	4.5	1359	9	US-09-739-254-13	Sequence 13, Appl
12	17	4.5	1666	10	US-09-938-842A-3427	Sequence 3427, Ap
13	17	4.5	1666	9	US-09-904-615-13	Sequence 13, Appl
14	17	4.5	1966	10	US-09-864-761-15087	Sequence 15087, A
15	17	4.5	6560	9	US-09-808-602-76	Sequence 76, Appl
16	17	4.5	9729	9	US-09-808-602-12	Sequence 12, Appl
17	17	4.5	9826	9	US-09-808-602-7	Sequence 7, Appl
18	17	4.5	40392	9	US-09-954-456-44	Sequence 44, Appl
19	17	4.5	40392	10	US-09-954-456-687	Sequence 687, App
20	17	4.5	659158	9	US-09-771-208-20	Sequence 20, Appl
21	16	4.2	191	10	US-09-878-574-10027	Sequence 10027, A
22	16	4.2	261	10	US-09-878-574-12914	Sequence 12914, A
23	16	4.2	275	10	US-09-923-876-513	Sequence 513, App
24	16	4.2	403	9	US-10-046-935-15	Sequence 15, Appl
25	16	4.2	403	10	US-09-878-178-15	Sequence 15, Appl
26	16	4.2	443	10	US-09-833-381-671	Sequence 671, App
27	16	4.2	465	10	US-09-833-381-1080	Sequence 1080, Ap
28	16	4.2	507	10	US-09-864-761-11919	Sequence 11919, A
29	16	4.2	551	10	US-09-864-761-14818	Sequence 14818, A
30	16	4.2	583	10	US-09-964-899-22	Sequence 22, Appl
31	16	4.2	636	10	US-09-938-842A-733	Sequence 733, App
32	16	4.2	869	10	US-09-770-445-565	Sequence 565, App
33	16	4.2	906	10	US-09-925-297-302	Sequence 302, App
34	16	4.2	936	10	US-09-886-055-354	Sequence 354, App
35	16	4.2	1006	9	US-09-964-899-22	Sequence 22, Appl
36	16	4.2	1212	9	US-09-938-842A-733	Sequence 733, App
37	16	4.2	1650	10	US-09-880-107-2199	Sequence 2199, Ap
38	16	4.2	1758	10	US-09-731-872-72	Sequence 72, Appl
39	16	4.2	1768	9	US-09-731-872-7	Sequence 7, Appl
40	16	4.2	1846	10	US-10-098-841-107	Sequence 107, App
41	16	4.2	1992	9	US-10-080-960-15	Sequence 15, Appl
42	16	4.2	2000	9	US-09-938-842A-3712	Sequence 3712, Ap
43	16	4.2	2003	10	US-09-887-576-306	Sequence 306, App
44	16	4.2	2004	10	US-09-887-576-253	Sequence 253, App
45	16	4.2	2481	10	US-09-852-909-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-864-761-7775
Sequence 7775, Application US20020048763A1

Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30

```
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;
;  SEQ ID NO 7775
;  LENGTH: 527
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  OTHER INFORMATION: MAP TO AL117341.26
;  OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL - 6.9
;  OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL - 6.3
;  OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL - 5.4
;  OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL - 6.4
;  OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL - 5.2
;  OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL - 6.4
;  OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL - 4.5
;  OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL - 5.7
;  OTHER INFORMATION: US-09-864-761-7775

Qy  162 TGTGTCTTTGGAATAAAAAC 180
        ||||||||||||||||||||
Db  477 TGTGTCTTTGGAATAAAAAC 495

Query Match       5.0%;    Score 19;    DB 10;    Length 527;
Best Local Similarity   100.0%;    Pred. No. 0.94;
Matches  19;   Conservative  0;   Mismatches  0;   Indels  0;   Gaps  0;

RESULT 2
US-09-910-943-287/c
Sequence 287, Application US/09910943
Patent No. US20020081610A1
GENERAL INFORMATION:
  APPLICANT: Hemmati-Brivanlou, Ali
  APPLICANT: Altman, Curtis
  TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression
  FILE REFERENCE: 7529/1G1480US1
  CURRENT APPLICATION NUMBER: US/09/910,943
  CURRENT FILING DATE: 2001-07-23
  NUMBER OF SEQ ID NOS: 742
  SOFTWARE: Patentin version 3.1
  SEQ ID NO 287
  LENGTH: 709
  TYPE: DNA
  ORGANISM: Xenopus laevis
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (1)..(709)
  OTHER INFORMATION: n may be a or g or c or t/u
US-09-910-943-287

Qy  259 GATGGTGATGGAGATGCCA 277
        |||||||||||||||||||
Db  255 GATGGTGATGGAGATGCCA 237

Query Match       5.0%;    Score 19;    DB 10;    Length 709;
Best Local Similarity   100.0%;    Pred. No. 0.95;
Matches  19;   Conservative  0;   Mismatches  0;   Indels  0;   Gaps  0;

RESULT 3
US-09-860-670-161
Sequence 161, Application US/09860670
Patent No. US20020165137A1
GENERAL INFORMATION:
  APPLICANT: Ruben et al.
```

```
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA127P1
;  CURRENT APPLICATION NUMBER: US/09,860,670
;  CURRENT FILING DATE: 2001-05-21
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 289
;  SOFTWARE: Patentin Ver. 2.0
;
;  SEQ ID NO 161
;  LENGTH: 26555
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  US-09-860-670-161

Qy  127 GTTAAGCCCTCGTTCTTT 144
        ||||||||||||||||||
Db  17975 GTTAAGCCCTCGTTCTTT 17992

Query Match       4.8%;    Score 18;    DB 9;    Length 26555;
Best Local Similarity   100.0%;    Pred. No. 3.8;
Matches  18;   Conservative  0;   Mismatches  0;   Indels  0;   Gaps  0;

RESULT 4
US-09-796-692-35/c
Sequence 35, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
  APPLICANT: Geiger, Alexander
  APPLICANT: Algate, Paul A.
  APPLICANT: Mannion, Jane
  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND T
  TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
  FILE REFERENCE: 2077.001200
  CURRENT APPLICATION NUMBER: US/09/796,692
  CURRENT FILING DATE: 2001-03-01
  PRIOR APPLICATION NUMBER: 60/186,126
  PRIOR FILING DATE: 2000-03-01
  PRIOR APPLICATION NUMBER: 60/190,479
  PRIOR FILING DATE: 2000-03-17
  PRIOR APPLICATION NUMBER: 60/200,545
  PRIOR FILING DATE: 2000-04-27
  PRIOR APPLICATION NUMBER: 60/200,303
  PRIOR FILING DATE: 2000-04-28
  PRIOR APPLICATION NUMBER: 60/200,779
  PRIOR FILING DATE: 2000-04-28
  PRIOR APPLICATION NUMBER: 60/200,999
  PRIOR FILING DATE: 2000-05-01
  PRIOR APPLICATION NUMBER: 60/202,084
  PRIOR FILING DATE: 2000-05-04
  PRIOR APPLICATION NUMBER: 60/206,201
  PRIOR FILING DATE: 2000-05-22
  PRIOR APPLICATION NUMBER: 60/218,950
  PRIOR FILING DATE: 2000-07-14
  PRIOR APPLICATION NUMBER: 60/222,903
  PRIOR FILING DATE: 2000-08-03
  PRIOR APPLICATION NUMBER: 60/223,416
  PRIOR FILING DATE: 2000-08-04
  PRIOR APPLICATION NUMBER: 60/223,378
  PRIOR FILING DATE: 2000-08-07
  NUMBER OF SEQ ID NOS: 9597
  SOFTWARE: FastSEQ for Windows Version 3.0
  SEQ ID NO 35
  LENGTH: 309
  TYPE: DNA
  ORGANISM: Homo sapiens
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (1)..(309)
  OTHER INFORMATION: n = A,T,C or G
US-09-796-692-35

Query Match       4.5%;    Score 17;    DB 9;    Length 309;
Best Local Similarity   100.0%;    Pred. No. 12;
```

RESULT 5
US-09-796-692-3533/c
Sequence 3533, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3533
; LENGTH: 309
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY:
; LOCATION: (11)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (229)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (272)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (277)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (289)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (296)
; OTHER INFORMATION: n=A,T,C or G
US-09-796-692-3533

Query Match   4.5%; Score 17; DB 9; Length 309;
Best Local Similarity   100.0%; Pred. No. 12;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   94 AATGTGTGTTCTGGAGC 110
          |||||||||||||||||
Db   68 AATGTGTGTTCTGGAGC 52

RESULT 6
US-09-878-574-3537
Sequence 3537, Application US/09878574
Patent No. US20020110548A1
GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 3537
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(397)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3028-009-Q1-B1-B5
US-09-878-574-3537

Query Match   4.5%; Score 17; DB 10; Length 397;
Best Local Similarity   100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   94 AATGTGTGTTCTGGAGC 110
          |||||||||||||||||
Db   68 AATGTGTGTTCTGGAGC 52

RESULT 7
US-10-040-739-843/c
Sequence 843, Application US/10040739
Patent No. US20020173635A1
GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
;   LaValle, Edward
;   Racie, Lisa
;   Merberg, David
;   Treacy, Maurice
;   Spaulding, Vikki
; TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
; NUMBER OF SEQUENCES: 1519
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genetics Institute, Inc.
;   STREET: 87 CambridgePark Drive
;   CITY: Cambridge
;   STATE: Massachusetts
;   COUNTRY: U.S.A
;   ZIP: 02140
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy Disk
;   COMPUTER: IBM PC Compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/040,739
;   FILING DATE: 07-Jan-2002
;   CLASSIFICATION: <Unknown>

Qy   103 TCTGGAGCTGAAGGATG 119
          |||||||||||||||||
Db   85 TCTGGAGCTGAAGGATG 101

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/036,520
FILING DATE: 03-JUN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 843:
SEQUENCE CHARACTERISTICS:
LENGTH: 627 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 843:
US-10-040-739-843

Qy   269 GAGATGCCAATCCATGG 285
         |||||||||||||||||
Db   610 GAGATGCCAATCCATGG 594

Query Match   4.5%;   Score 17;   DB 9;   Length 627;
Best Local Similarity   100.0%;
Matches   17;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 8
US-09-938-842A-1704
Sequence 1704, Application US/0993842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
SAME, AND METHODS OF USE
FILE REFERENCE: SCRIP1300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1704
LENGTH: 837
ORGANISM: Arabidopsis thaliana
TYPE: DNA
US-09-938-842A-1704

Qy   229 TGAAAAGAGACCTTTGG 245
         |||||||||||||||||
Db   276 TGAAAAGAGACCTTTGG 292

Query Match   4.5%;   Score 17;   DB 9;   Length 837;
Best Local Similarity   100.0%;
Matches   17;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 9
US-09-867-550-1051/c
Sequence 1051, Application US/09867550
Patent No. US20020082206A1
GENERAL INFORMATION:
APPLICANT: Leach, Martin D.
APPLICANT: Mehraban, Fuad
APPLICANT: Conley, Pamela
APPLICANT: Law, Debbie
APPLICANT: Topper, James
TITLE OF INVENTION: No. US20020082206A1e1 Polynucleotides from Atherogenic Cell
TITLE OF INVENTION: Thereby
FILE REFERENCE: 21402-013 (Cura-313)
CURRENT APPLICATION NUMBER: US/09/867,550
PRIOR APPLICATION NUMBER: USSN 60/208,427
PRIOR FILING DATE: 2000-05-30
NUMBER OF SEQ ID NOS: 2125
SOFTWARE: FastSEQ for Windows Version 4.0
OTHER INFORMATION: wherein any n is one of a or t or g or c
SEQ ID NO 1051
LENGTH: 1005
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(2)
OTHER INFORMATION: wherein any n is one of a or t or g or c
US-09-867-550-1051

Qy   271 GATGCCAATCCATGGAA 287
         |||||||||||||||||
Db   600 GATGCCAATCCATGGAA 584

Query Match   4.5%;   Score 17;   DB 10;   Length 1005;
Best Local Similarity   100.0%;
Matches   17;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 10
US-09-864-761-31616/c
Sequence 31616, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 31616
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004526.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.9
; OTHER INFORMATION: SWISSPROT HIT: P34910, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P34910, EVALUE 0.00e+00
; OTHER INFORMATION: NT HIT: gi11426548, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AA805896.1, EVALUE 0.00e+00
US-09-864-761-31616

Qy 301 CTATGTTGGTAGCTATA 317
         |||||||||||||||||
Db 467 CTATGTTGGTAGCTATA 451

Query Match   4.5%; Score 17; DB 10; Length 1041;
Best Local Similarity   100.0%; Pred. No. 12;
Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 11
US-09-938-842A-3427/c
Sequence 3427, Application US20020160378A1
; Patent No. US20020160378A1
GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3427
; TYPE: DNA
; LENGTH: 1359
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3427

Qy   76 GAGTCTCTCACAAGCTT 92
         ||||||||||||||||
Db 531 GAGTCTCTCACAAGCTT 515

Query Match   4.5%; Score 17; DB 9; Length 1359;
Best Local Similarity   100.0%; Pred. No. 12;
Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 12
US-09-739-254-13/c
; Sequence 13, Application US/0973925A
; Patent No. US20010021700A1

GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: PZ032P1
; CURRENT APPLICATION NUMBER: US/09/739,254
; CURRENT FILING DATE: 2000-12-19
; EARLIER APPLICATION NUMBER: 09/511,554
; EARLIER FILING DATE: 2000-02-23
; EARLIER APPLICATION NUMBER: PCT/US99/19330
; EARLIER FILING DATE: 1999-08-24
; EARLIER APPLICATION NUMBER: 60/097,917
; EARLIER FILING DATE: 1998-08-25
; EARLIER APPLICATION NUMBER: 60/098,634
; EARLIER FILING DATE: 1998-08-31
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 13
; TYPE: DNA
; LENGTH: 1666
; ORGANISM: Homo sapiens
US-09-739-254-13

Qy 271 GATGCCAATCCATGGAA 287
         ||||||||||||||||
Db 740 GATGCCAATCCATGGAA 724

Query Match   4.5%; Score 17; DB 10; Length 1666;
Best Local Similarity   100.0%; Pred. No. 12;
Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 13
US-09-904-615-13/c
; Sequence 13, Application US/0904615
; Patent No. US20020026040A1
GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: PZ032P1
; CURRENT APPLICATION NUMBER: US/09/904,615
; CURRENT FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/511,554
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/097,917
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 60/098,634
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 13
; TYPE: DNA
; LENGTH: 1666
; ORGANISM: Homo sapiens
US-09-904-615-13

Qy 271 GATGCCAATCCATGGAA 287
         ||||||||||||||||
Db 740 GATGCCAATCCATGGAA 724

Query Match   4.5%; Score 17; DB 10; Length 1666;
Best Local Similarity   100.0%; Pred. No. 12;
Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 14
US-09-864-761-15087/c
; Sequence 15087, Application US/09864761
; Patent No. US20020048763A1
GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng

```
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
NUMBER OF SEQ ID NOS: 4917
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 15087
LENGTH: 1966
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004526.1
OTHER INFORMATION: EXPRESSED IN HB1100, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.9
```

```
RESULT 15
US-09-808-602-76/c
Sequence 76, Application US/09808602
Patent No. US20020155115A1
GENERAL INFORMATION:
APPLICANT: Vernet, Corine A
APPLICANT: Fernandes, Elma
APPLICANT: Shimkets, Richard A
APPLICANT: Herrman, John L
APPLICANT: Majunder, Kumud
APPLICANT: Mishra, Vishnu
APPLICANT: Mezes, Peter S
APPLICANT: MacDougall, John
TITLE OF INVENTION: NO. US20020155115A1e1 proteins and Nuclec Acids Encoding Sa
FILE REFERENCE: 15966-697 CIP
CURRENT APPLICATION NUMBER: US/09/808,602
CURRENT FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 09/800,198
PRIOR FILING DATE: 2001-03-05
PRIOR APPLICATION NUMBER: 60/186,596
PRIOR FILING DATE: 2000-03-03
NUMBER OF SEQ ID NOS: 114
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 76
LENGTH: 6560
TYPE: DNA
ORGANISM: Homo sapiens
US-09-808-602-76
```

```
Query Match             4.5%;  Score 17;  DB 10;  Length 1966;
Best Local Similarity   100.0%;
Matches    17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    301 CTATGTTGGTAGCTATA 317
          |||||||||||||||||
Db    810 CTATGTTGGTAGCTATA 794
```

```
Query Match             4.5%;  Score 17;  DB 9;  Length 6560;
Best Local Similarity   100.0%;
Matches    17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    185 CTGCCAAGTAGTACTTG 201
          |||||||||||||||||
Db    767 CTGCCAAGTAGTACTTG 751

Search completed: January 31, 2003, 05:19:41
Job time : 345 secs
```

GenCore version 5.1.3

Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2003, 01:19:33 ; Search time 53 Seconds
                                         (without alignments)
                                         2187.242 Million cell updates/sec

Title: US-09-803-719-222
Perfect score: 378
Sequence: 1 tacggctgcttaagacgact..............taccattggagaaccatgcg 378

Scoring table: OLIGO_NUC
                Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 15338381 residues

Word size :   0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :   Issued_Patents_NA:*
                1: /cgn2_6/ptodata/1/1na/5A_COMB.seq:*
                2: /cgn2_6/ptodata/1/1na/5B_COMB.seq:*
                3: /cgn2_6/ptodata/1/1na/6A_COMB.seq:*
                4: /cgn2_6/ptodata/1/1na/6B_COMB.seq:*
                5: /cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:*
                6: /cgn2_6/ptodata/1/1na/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	17	4.5	1080	4	US-09-134-001C-1264	Sequence 1264, Ap
2	16	4.2	237	1	US-08-480-910-6	Sequence 6, Appli
3	16	4.2	1146	4	US-09-257-580-1	Sequence 1, Appli
4	16	4.2	1146	4	US-09-257-580-4	Sequence 4, Appli
5	16	4.2	1285	5	US-08-480-910-11	Sequence 11, Appli
6	16	4.2	1285	5	PCT-US95-00052-1	Sequence 1, Appli
7	16	4.2	1314	4	US-09-500-569-17	Sequence 17, Appli
8	16	4.2	1873	4	US-09-461-474-9	Sequence 9, Appli
9	16	4.2	4413	2	US-09-221-017B-811	Sequence 811, App
10	16	4.2	18596	4	US-09-318-448-11	Sequence 11, Appl
11	16	4.2	34303	2	US-08-735-609-4	Sequence 4, Appli
12	16	4.2	34303	2	US-08-735-609-4	Sequence 4, Appli
13	16	4.2	34303	3	US-09-315-372-4	Sequence 4, Appli
14	16	4.2	34303	3	US-09-244-752-4	Sequence 4, Appli
15	16	4.2	34303	3	US-09-245-497-4	Sequence 4, Appli
16	16	4.2	34303	3	US-09-562-919-4	Sequence 4, Appli
17	16	4.2	34382	2	US-08-374-483-6	Sequence 6, Appli
18	16	4.2	35408	4	US-08-973-334-3	Sequence 3, Appli
19	16	4.2	35408	4	US-09-563-869A-3	Sequence 3, Appli
20	16	4.2	35408	4	US-08-549-489-3	Sequence 3, Appli
21	16	4.2	35935	2	US-08-735-609-1	Sequence 1, Appli
22	16	4.2	35935	2	US-08-379-452-43	Sequence 43, Appl
23	16	4.2	35935	3	US-09-315-372-1	Sequence 1, Appli
24	16	4.2	35935	3	US-09-244-752-1	Sequence 1, Appli
25	16	4.2	35935	3	US-09-245-497-1	Sequence 1, Appli
26	16	4.2	35935	3	US-09-245-497-1	Sequence 1, Appli
27	16	4.2	35935	3	US-09-409-670-43	Sequence 43, Appli
28	16	4.2	35935	4	US-09-562-919-1	Sequence 1, Appli
29	16	4.2	48974	4	US-08-920-422-17	Sequence 17, Appl
30	16	4.2	246240	2	US-08-724-394A-20	Sequence 20, Appl
31	16	4.2	246240	2	US-08-724-394A-21	Sequence 21, Appl
32	16	4.2	246240	2	US-08-724-394A-22	Sequence 22, Appl
33	15	4.0	22	2	US-08-734-941-2	Sequence 2, Appli
34	15	4.0	32	4	US-09-923-992A-29	Sequence 29, Appli
35	15	4.0	49	4	US-09-390-867A-25	Sequence 25, Appli
36	15	4.0	49	4	US-08-548-260-25	Sequence 25, Appli
37	15	4.0	81	4	US-08-895-590-23	Sequence 23, Appli
38	15	4.0	89	4	US-09-390-867A-23	Sequence 23, Appli
39	15	4.0	89	4	US-08-548-260-23	Sequence 23, Appli
40	15	4.0	280	4	US-09-026-343-31	Sequence 31, Appli
41	15	4.0	280	4	US-08-362-871-31	Sequence 31, Appli
42	15	4.0	344	1	US-08-644-426-3	Sequence 3, Appli
43	15	4.0	444	4	US-07-925-978B-1	Sequence 1, Appli
44	15	4.0	451	4	US-09-385-982-55	Sequence 55, Appli
45	15	4.0	540	3	US-09-042-771-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-134-001C-1264

; Sequence 1264, Application US/09134001C

; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCO
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE:
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1264
;   LENGTH: 1080
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1264

Query Match   4.5%; Score 17; DB 4; Length 1080;
Best Local Similarity   100.0%; Pred. No. 14;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   309 GTAGCTATAGCAGAGAAGT 325
            |||||||||||||||||||
Db   889 GTAGCTATAGCAGAGAAGT 905

RESULT 2
US-08-480-910-6

; Sequence 6, Application US/08480910
; Patent No. 5693530
; GENERAL INFORMATION:
; APPLICANT: Karel A. Schat, Kazuhiko Ohashi, and Priscilla
; APPLICANT: H. O'Connell
; TITLE OF INVENTION: A Marek's Disease Virus Nucleotide
; TITLE OF INVENTION: Sequence and Methods of Use
; Patent NO. 5693530
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
;   STREET: 1800 One M&T Plaza
;   CITY: Buffalo
;   STATE: New York
;   COUNTRY: United States
;   ZIP: 14203-2391
; COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44Mb storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS/ Microsoft Windows
SOFTWARE: Wordperfect for Windows 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,910
FILING DATE: 07 June 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. Serial No. 5693530 08/180,051
FILING DATE: 11 January 1994
ATTORNEY/AGENT INFORMATION:
NAME: Nelson, M. Bud
REGISTRATION NUMBER: 35,300
REFERENCE/DOCKET NUMBER: 18617.0005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 856-4000
TELEFAX: (716) 849-0349
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-480-910-6

Qy   49 AGGAACATATGGAAAA 64
         ||||||||||||||||
Db   34 AGGAACATATGGAAAA 49

Query Match   4.2%; Score 16; DB 1; Length 237;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3
US-09-257-580-1
; Sequence 1, Application US/09257580
; Patent No. 6307036
; GENERAL INFORMATION:
; APPLICANT: Yorkshire Cancer Research
; TITLE OF INVENTION: Tumour Suppressor Gene
; FILE REFERENCE: Canine p53
; CURRENT APPLICATION NUMBER: US/09/257,580
; CURRENT FILING DATE: 1999-02-25
; PRIOR APPLICATION NUMBER: 9804178.3
; PRIOR FILING DATE: 1998-02-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 1
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: Canis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1146)
US-09-257-580-1

Qy   105 TGGAGCTGAAGGATGC 120
         ||||||||||||||||
Db 1007 TGGAGCTGAAGGATGC 1022

RESULT 4
US-09-257-580-4
; Sequence 4, Application US/09257580
; Patent No. 6307036
; GENERAL INFORMATION:
; APPLICANT: Yorkshire Cancer Research

TITLE OF INVENTION: Tumour Suppressor Gene
FILE REFERENCE: Canine p53
CURRENT APPLICATION NUMBER: US/09/257,580
CURRENT FILING DATE: 1999-02-25
PRIOR APPLICATION NUMBER: 9804178.3
PRIOR FILING DATE: 1998-02-28
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 1146
TYPE: RNA
ORGANISM: canis
US-09-257-580-4

Qy   105 TGGAGCTGAAGGATGC 120
         ||||||||||||||||
Db 1007 UGGAGCUGAAGGAUGC 1022

Query Match   4.2%; Score 16; DB 4; Length 1146;
Best Local Similarity 81.2%; Pred. No. 44;
Matches 13; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

RESULT 5
US-08-480-910-1
; Sequence 1, Application US/08480910
; Patent No. 5693530
; GENERAL INFORMATION:
; APPLICANT: Karel A. Schat, Kazuhiko Ohashi, and Priscilla
; APPLICANT: H. O'Connell
; TITLE OF INVENTION: A Marek's Disease Virus Nucleotide
; TITLE OF INVENTION: Sequence and Methods of Use
; Patent No. 5693530
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
; STREET: 1800 One M&T Plaza
; CITY: Buffalo
; STATE: New York
; COUNTRY: United States
; ZIP: 14203-2391
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44Mb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS/ Microsoft Windows
; SOFTWARE: Wordperfect for Windows 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,910
; FILING DATE: 07 June 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. Serial No. 5693530 08/180,051
; FILING DATE: 11 January 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Nelson, M. Bud
; REGISTRATION NUMBER: 35,300
; REFERENCE/DOCKET NUMBER: 18617.0005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 856-4000
; TELEFAX: (716) 849-0349
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1285 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single-stranded
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: yes
; IMMEDIATE SOURCE:
; LIBRARY: cDNA
; CLONE: clone L1
; ORIGINAL SOURCE:
; ORGANISM: Marek's Disease Virus
; STRAIN: MDCC-CU41

CELL TYPE: Virus
FEATURE:
LOCATION: L1 open reading frame, 725-1045
IDENTIFICATION METHOD: by experiment
OTHER INFORMATION:
US-08-480-910-1

Qy   49 AGGAACATATGGAAAA 64
            ||||||||||||||||
Db   761 AGGAACATATGGAAAA 776

Query Match   4.2%;   Score 16;   DB 1;   Length 1285;
Best Local Similarity 100.0%;   Pred. No. 44;
Matches 16;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

RESULT 6
PCT-US95-00052-1
Sequence 1, Application PC/TUS9500052
GENERAL INFORMATION:
APPLICANT: Karel A. Schat, Kazuhiko Ohashi, and
APPLICANT: Priscilla H. O'Connell
TITLE OF INVENTION: Control Of Marek's Disease By The
TITLE OF INVENTION: Inhibition Of Latency And Tumor Cell Development
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hodgson, Russ, Andrews, Woods &
ADDRESSEE: Goodyear
STREET: 1800 One M&T Plaza
CITY: Buffalo
STATE: New York
COUNTRY: United States
ZIP: 14203-2391
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1
SOFTWARE: Wordperfect for Windows 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00052
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. Serial No. 08/180,051
FILING DATE: January 11, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Nelson, M. Bud
REGISTRATION NUMBER: 35,300
REFERENCE/DOCKET NUMBER: 18617.0005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 856-4000
TELEFAX: (716) 849-0349
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1285 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: yes
IMMEDIATE SOURCE:
LIBRARY: cDNA
CLONE: clone L1
ORIGINAL SOURCE:
ORGANISM: Marek's Disease Virus
STRAIN: MDCC-CU41
CELL TYPE: virus
FEATURE:
LOCATION: L1 open reading frame, 725-1045
IDENTIFICATION METHOD: by experiment
OTHER INFORMATION:
PCT-US95-00052-1

Query Match   4.2%;   Score 16;   DB 5;   Length 1285;

Best Local Similarity 100.0%;   Pred. No. 44;
Matches 16;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   49 AGGAACATATGGAAAA 64
            ||||||||||||||||
Db   761 AGGAACATATGGAAAA 776

RESULT 7
US-09-500-569-17/c
Sequence 17, Application US/09500569
Patent No. 6329204
GENERAL INFORMATION:
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Rafalski, Antoni
APPLICANT: Shen, Jennie
TITLE OF INVENTION: Plant Caffeic acid 3-O-Methyltransferase Homologs
FILE REFERENCE: BB1327 US NA
CURRENT APPLICATION NUMBER: US/09/500,569
CURRENT FILING DATE: 2000-02-09
EARLIER APPLICATION NUMBER: 60/119,587
EARLIER FILING DATE: 1999-February-10
NUMBER OF SEQ ID NOS: 28
SOFTWARE: Microsoft Office 97
SEQ ID NO 17
LENGTH: 1314
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (472)
FEATURE:
NAME/KEY: unsure
LOCATION: (1156)
FEATURE:
NAME/KEY: unsure
LOCATION: (1180)
FEATURE:
NAME/KEY: unsure
LOCATION: (1262)
FEATURE:
NAME/KEY: unsure
LOCATION: (1302)
US-09-500-569-17

Qy   49 AGGAACATATGGAAAA 64
            ||||||||||||||||
Db   761 AGGAACATATGGAAAA 776

Query Match   4.2%;   Score 16;   DB 4;   Length 1314;
Best Local Similarity 100.0%;   Pred. No. 44;
Matches 16;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

RESULT 8
US-09-461-474-9/c
Sequence 9, Application US/09461474
Patent No. 6278042
GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Rafalski, Antoni
APPLICANT: Sakai, Hajime
TITLE OF INVENTION: Plant Metal Transporters
FILE REFERENCE: BB1303 US NA
CURRENT APPLICATION NUMBER: US/09/461,474
CURRENT FILING DATE: 1999-12-14
EARLIER APPLICATION NUMBER: 60/112,562
EARLIER FILING DATE: 1998-12-16
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Microsoft Office 97
SEQ ID NO 9
LENGTH: 1873
TYPE: DNA

Qy   67 CCCAAGCCGGAGTCTC 82
            ||||||||||||||||
Db   1247 CCCAAGCCGGAGTCTC 1232

; ORGANISM: Oryza sativa

US-09-461-474-9
Query Match
Best Local Similarity 100.0%; Pred. No. 43;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;   Score 16; DB 4; Length 1873;

Qy   258 GGATGGTGATGGAGAT 273
         ||||||||||||||||
Db   885 GGATGGTGATGGAGAT 870

RESULT 9
US-09-221-017B-811/c
Sequence 811, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 811:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4413 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..4413

US-09-221-017B-811

Query Match 4.2%; Score 16; DB 4; Length 4413;

Best Local Similarity 100.0%; Pred. No. 41;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   229 TGAAAAGAGACCTTTG 244
         ||||||||||||||||
Db   845 TGAAAAGAGACCTTTG 830

RESULT 10
US-09-318-448-11
Sequence 11, Application US/09318448
Patent No. 6210950
GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 18596
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-448-11

Qy   .259 GATGGTGATGGAGATG 274
          ||||||||||||||||
Db 10134 GATGGTGATGGAGATG 10149

Query Match 4.2%; Score 16; DB 4; Length 18596;

RESULT 11
US-08-735-609-4/c
Sequence 4, Application US/08735609
Patent No. 5955360
GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amalfitano, Andrea
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Hartigan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/735,609
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02484
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 4:

RESULT 12
US-08-735-609-4/c
Sequence 4, Application US/08735609
Patent No. 5994132
GENERAL INFORMATION:
    APPLICANT: Chamberlain, Jeffrey S.
    APPLICANT: Amalfitano, Andrea
    APPLICANT: Hauser, Michael A.
    APPLICANT: Kumar-Singh, Rajendra
    APPLICANT: Hartigan-O'Connor, Dennis J.
    TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
    NUMBER OF SEQUENCES: 15
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Medlen & Carroll, LLP
        STREET: 220 Montgomery Street, Suite 2200
        CITY: San Francisco
        STATE: California
        COUNTRY: United States of America
        ZIP: 94104
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: Patentin Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/735,609
        FILING DATE: 23-Oct-1996
        CLASSIFICATION: <Unknown>
    ATTORNEY/AGENT INFORMATION:
        NAME: Ingolia, Diane E.
        REGISTRATION NUMBER: 40,027
        REFERENCE/DOCKET NUMBER: UM-02484
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: (415) 705-8410
        TELEFAX: (415) 397-8338
    INFORMATION FOR SEQ ID NO: 4:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 34303 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: double
        TOPOLOGY: linear
        MOLECULE TYPE: other nucleic acid
        DESCRIPTION: /desc = "DNA"
    SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Qy   212 CAACATGCTTTGACTG 227
          ||||||||||||||||
Db 27052 CAACATGCTTTGACTG 27037

Query Match   4.2%;   Score 16;   DB 2;   Length 34303;
Best Local Similarity   100.0%;   Pred. No. 37;
Matches   16;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

US-08-735-609-4

Qy   212 CAACATGCTTTGACTG 227
          ||||||||||||||||
Db 27052 CAACATGCTTTGACTG 27037

; SEQUENCE CHARACTERISTICS:
;   LENGTH: 34303 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: other nucleic acid
;   DESCRIPTION: /desc = "DNA"

RESULT 13
US-09-315-372-4/c
Sequence 4, Application US/09315372
Patent No. 6057158
GENERAL INFORMATION:
    APPLICANT: Chamberlain, Jeffrey S.
    APPLICANT: Amalfitano, Andrea
    APPLICANT: Hauser, Michael A.
    APPLICANT: Kumar-Singh, Rajendra
    APPLICANT: Hartigan-O'Connor, Dennis J.
    TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
    NUMBER OF SEQUENCES: 15
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Medlen & Carroll, LLP
        STREET: 220 Montgomery Street, Suite 2200
        CITY: San Francisco
        STATE: California
        COUNTRY: United States of America
        ZIP: 94104
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: Patentin Release #1.0, Version #1.30
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 08/735,609
        FILING DATE:
    CLASSIFICATION:
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/09/315,372
        FILING DATE:
    ATTORNEY/AGENT INFORMATION:
        NAME: Ingolia, Diane E.
        REGISTRATION NUMBER: 40,027
        REFERENCE/DOCKET NUMBER: UM-02484
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: (415) 705-8410
        TELEFAX: (415) 397-8338
    INFORMATION FOR SEQ ID NO: 4:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 34303 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: double
        TOPOLOGY: linear
        MOLECULE TYPE: other nucleic acid
        DESCRIPTION: /desc = "DNA"
    SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Qy   212 CAACATGCTTTGACTG 227
          ||||||||||||||||
Db 27052 CAACATGCTTTGACTG 27037

Query Match   4.2%;   Score 16;   DB 3;   Length 34303;
Best Local Similarity   100.0%;   Pred. No. 37;
Matches   16;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

US-09-315-372-4

RESULT 14
US-09-244-752-4/c
Sequence 4, Application US/09244752
Patent No. 6063622
GENERAL INFORMATION:
    APPLICANT: Chamberlain, Jeffrey S.
    APPLICANT: Amalfitano, Andrea
    APPLICANT: Hauser, Michael A.
    APPLICANT: Kumar-Singh, Rajendra
    APPLICANT: Hartigan-O'Connor, Dennis J.
    TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
    NUMBER OF SEQUENCES: 15
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Medlen & Carroll, LLP
        STREET: 220 Montgomery Street, Suite 2200
        CITY: San Francisco

; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02484
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
; US-09-245-497-4

Query Match   4.2%;   Score 16; DB 3; Length 34303;
Best Local Similarity 100.0%; Pred. No. 37;
Matches   16; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   212 CAACATGCTTTGACTG 227
             ||||||||||||||||
Db   27052 CAACATGCTTTGACTG 27037

Search completed: January 31, 2003, 03:51:24
Job time : 133 secs

STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,752
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/735,609
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02484
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 34303 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-244-752-4

Query Match   4.2%;   Score 16; DB 3; Length 34303;
Best Local Similarity 100.0%; Pred. No. 37;
Matches   16; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   212 CAACATGCTTTGACTG 227
             ||||||||||||||||
Db   27052 CAACATGCTTTGACTG 27037

RESULT 15
US-09-245-497-4/c
Sequence 4, Application US/09245497
Patent No. 6083750
GENERAL INFORMATION:
APPLICANT: Chamberlain, Jeffrey S.
APPLICANT: Amalfitano, Andrea
APPLICANT: Hauser, Michael A.
APPLICANT: Kumar-Singh, Rajendra
APPLICANT: Hartigan-O'Connor, Dennis J.
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/245,497
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/735,609
APPLICATION NUMBER: <B> FILING DATE:
ATTORNEY/AGENT INFORMATION:

```
GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:           January 31, 2003, 03:22:50 ; Search time 2218 Seconds
                  (without alignments)
                  2760.096 Million cell updates/sec

Title:            US-09-803-719-222
Perfect score:    378
Sequence:         1 tacggctgcttaagacgact.........taccattggagaaccatgcg 378

Scoring table:    OLIGO_NUC
                  Gapop 60.0 , Gapext 60.0
Searched:         16154066 seqs, 8097743376 residues
Word size :       0

Total number of hits satisfying chosen parameters:      32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries
Database :        EST:*

                  1: em_estba:*
                  2: em_esthum:*
                  3: em_estin:*
                  4: em_estmu:*
                  5: em_estov:*
                  6: em_estpl:*
                  7: em_estro:*
                  8: em_htc:*
                  9: gb_est1:*
                 10: gb_est2:*
                 11: gb_htc:*
                 12: gb_est3:*
                 13: gb_est4:*
                 14: gb_est5:*
                 15: em_estfun:*
                 16: em_estom:*
                 17: gb_gss:*
                 18: em_gss_hum:*
                 19: em_gss_inv:*
                 20: em_gss_pln:*
                 21: em_gss_vrt:*
                 22: em_gss_fun:*
                 23: em_gss_mam:*
                 24: em_gss_mus:*
                 25: em_gss_other:*
                 26: em_gss_pro:*
                 27: em_gss_rod:*
```

SUMMARIES

```
Result                        %         Query
No.   Score   Match   Length  DB ID                     Description

  1    90     23.8    451     BM987874    BM987874 UI-H-COO-
  2    49     13.0    386     AA743908    AA743908 ob05b10.s
  3    49     13.0    439     AI459918    AI459918 ar81h09.x
  4    45     11.9    806     BG197212    BG197212 RST16451
  5    34      9.0    349     AQ035618    AQ035618 CIT-HSP-2
  6    30      7.9    237     AQ035738    AQ035738 CIT-HSP-2
```

```
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

```
c   7   21  5.6  357  12  BG210731   BG210731 RST30277
c   8   21  5.6  440  14  BQ597872   BQ597872 MI-P-A2-a
c   9   21  5.6  526  12  BG223756   BG223756 IM0018F1
c  10   21  5.6  541  10  AV605455   AV605455 AV605455
c  11   21  5.6  545  10  AV605117   AV605117 AV605117
c  12   21  5.6  547  12  BF713089   BF713089 MI-P-H3-a
c  13   21  5.6  569  10  AW464207   AW464207 BP230015A
c  14   21  5.6  589  17  A2364217   A2364217 IM010P18
c  15   21  5.6  611  14  BQ600629   BQ600629 MI-P-E7-a
c  16   21  5.6  628  14  BQ599839   BQ599839 MI-P-E6-a
c  17   21  5.6  635  10  BQ597758   BQ597758 MI-P-CP1-
c  18   21  5.6  666  13  AV610564   AV610564 MI-P-A2-a
c  19   21  5.6  680  13  BI186343   BI186343 UNL-P-FN-
c  20   21  5.6  692  14  BQ604599   BQ604599 MI-P-FN-
c  21   21  5.6  785  13  BI184120   BI184120 UNL-P-FN-
c  22   20  5.3  144  12  BE769032   BE769032 PM4-FT002
c  23   20  5.3  379  13  BE680526   BE680526 df81d04.y
c  24   20  5.3  391  10  AL672699   AL672699 AL672699
c  25   20  5.3  440  17  AQ204259   AQ204259 HS_3116_B
c  26   20  5.3  507   9  AL638747   AL638747 AL638747
c  27   20  5.3  542   9  AL798424   AL798424 AL798424
c  28   20  5.3  640  13  BJ094175   BJ094175 BJ094175
c  29   20  5.3  769  17  AQ916538   AQ916538 nbeb0065M
c  30   20  5.3  812  17  BH703215   BH703215 BOMAQ70TF
c  31   20  5.3  847  14  BQ885975   BQ885975 AGENCOURT
c  32   19  5.0  310  12  BF760676   BF760676 CM1-CT063
c  33   19  5.0  334  13  BM443382   BM443382 EbroC02_SQ
c  34   19  5.0  347  17  A2651179   A2651179 1M052SJ15
c  35   19  5.0  356  14  C28465     C28465 C28465 Rice
c  36   19  5.0  363  12  BF362917   BF362917 PM4-NN009
c  37   19  5.0  369  17  AQ524481   AQ524481 HS_5214_B
c  38   19  5.0  436  14  BQ762497   BQ762497 EbroC02_SQ
c  39   19  5.0  506  17  AQ441231   AQ441231 HS_5129_B
c  40   19  5.0  523   9  AL697366   AL697366 AL697366
c  41   19  5.0  534  17  AF175356   AF175356 AF175356
c  42   19  5.0  574  13  BM657418   BM657418 170006874
c  43   19  5.0  578  14  BQ579887   BQ579887 AC806-Jm7
c  44   19  5.0  616  12  BG409744   BG409744 S10-4-A1
c  45   19  5.0  631  13  BM635028   BM635028 170006875
```

ALIGNMENTS

```
RESULT 1
BM987874/c
LOCUS       BM987874                   451 bp    mRNA    linear    EST 17-JUN-2002
DEFINITION  UI-H-COO-asz-a-11-0-UI.s1 NCI_CGAP_Sub9 Homo sapiens cDNA clone
            IMAGE:5860149 3', mRNA sequence.
ACCESSION   BM987874
VERSION     BM987874.1  GI:19707263
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 451)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Seq primer: M13 FORWARD
            PolyA=Yes.
FEATURES             Location/Qualifiers
     source          1..451
```

```
FEATURES             Location/Qualifiers
     source          1..386
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5860149"
                     /clone_lib="NCI_CGAP_Sub9"
                     /lab_host="DH10B"
                     /dev_stage="mixed"
                     /tissue_type="mixed"
                     /note="Vector: pT7T3-Pac (Life Technologies)" with a modified
                     polylinker; Site_1: EcoR I; Site_2: Not I; tissues:
                     adenosquamous carcinoma, Ligament cells, Prostate
                     carcinoma, Bladder carcinoma, Brain oligodenroga ;
                     ulcerative colitis, Fetal thymus, Cervix, Cervical
                     Cholonic mucosa with Crohns disease, Cholonic mucosa with
                     NCI_CGAP_Sub9 is a subtracted cDNA library constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tags for this library are CGTC, AACG, GGCC,
                     GGAAG, TAGC, TAAGC, ATGG, AGACA, ATCAC. For additional
                     information, contact: Bento Soares, bento-soares@uiowa.edu
                     TAG_LIB=UI-H-C00
                     TAG_TISSUE=Prostate Carcinoma
                     TAG_SEQ=ATGG"
BASE COUNT      109 a    93 c    83 g    166 t
ORIGIN

Query Match      23.8%;  Score 90;  Pred. No. 2.4e-39;  DB 14;  Length 451;
Best Local Similarity  100.0%;
Matches  90;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   207 ATGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGCACACACATTGAAGGGATGGTGA 266
         ||||||||||||||||||||||||||||||||||||||| ||||| ||||||||||||||||
Db   451 ATGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACANTTGAAGGGATGGTGA 392

Qy   267 TGGAGATGCCAAATCCATGGAAATCAGGTGCC 296
         |||||||||||||||||||||||||||||||||
Db   391 TGGAGATGCCAAATCCATGGAAATCAGGTGCC 362
```

```
RESULT 2
AA743908/c
LOCUS        AA743908        386 bp    mRNA    linear    EST 19-FEB-1998
DEFINITION   ob05b10.s1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1322779 3',
             mRNA sequence.
ACCESSION    AA743908
VERSION      AA743908.1  GI:2784658
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 386)
  AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@email.nih.gov
             unknown library type
             Insert Length: 508   Std Error: 0.00
             Seq primer: -40m13 fwd. ET from Amersham
             High quality sequence stop: 234.
FEATURES             Location/Qualifiers
     source          1..386
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1322779"
                     /clone_lib="NCI_CGAP_Kid3"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer,
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. mRNA
                     source: 2 pooled kidneys. Library went through one round
                     of normalization. Library constructed by Bento Soares and
                     M. Fatima Bonaldo."
BASE COUNT      93 a    77 c    73 g    143 t
ORIGIN

Query Match      13.0%;  Score 49;  Pred. No. 2.9e-16;  DB 9;  Length 386;
Best Local Similarity  100.0%;
Matches  49;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   298 CAGCTATGTGGTAGCTATAGCAGAAGTGTCTTCTTGGCAAAGATTCCTCC 346
         |||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   354 CAGCTATGTTGGGTAGCTATAGCAGAAGTGTCTTCTTGGCAAAGATTCCTCC 306
```

```
RESULT 3
AI459918/c
LOCUS        AI459918        439 bp    mRNA    linear    EST 09-MAR-1999
DEFINITION   ar81h09.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone
             IMAGE:2151713 3', mRNA sequence.
ACCESSION    AI459918
VERSION      AI459918.1  GI:4312799
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 439)
  AUTHORS    Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
             Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
             ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
             White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE      Washu-NCI human EST Project
  JOURNAL    Unpublished (1997)
COMMENT      Contact: Wilson RK
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: est@watson.wustl.edu
             This clone is available on a royalty-free through LLNL ; contact
             the IMAGE Consortium (info@image.llnl.gov) for further information.
             Seq primer: -40UP from Gibco
             High quality sequence stop: 425.
FEATURES             Location/Qualifiers
     source          1..439
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2151713"
                     /clone_lib="Barstead colon HPLRB7"
                     /sex="male"
                     /dev_stage="adult, age 25"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTAGAGTCGGCGCGCCGCGCGCCCTTTTTTTTTTTTTTTTTTTTT
                     3'], double-stranded cDNA was ligated to Eco RI adaptors
                     [5' AATTCACTAGTAAT 3' and 5' ATTACTAGTG 3'], digested
                     with Not I and cloned into the Not I and Eco RI sites of
                     the modified pT7T3 vector. Library constructed by Bob
                     Barstead."
BASE COUNT      115 a    89 c    81 g    154 t
ORIGIN

Query Match      13.0%;  Score 49;  DB 9;  Length 439;
```

Best Local Similarity 100.0%; Pred. No. 2.9e-16;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   298 CAGCTATGTTGGTAGCTATAGCAGAAGAAGTCTTCTTGGCAAAGATTCCTCC 346
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   397 CAGCTAGTTGGTAGCTATAGCAGAAGAAGTCTTCTTGGCAAAGATTCCTCC 349

RESULT 4
BG197212/c
LOCUS   BG197212   806 bp   mRNA   linear   EST 21-APR-2001
DEFINITION   RST16451 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG197212
VERSION   BG197212.1   GI:13718899
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 806)
AUTHORS   Harrington,J.J., Sherf,B., Rundlett,S., Perry,R.,
              Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.
              , Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith
              ,E., Veloso,N., Kilka,A., Hess,J., Cothren,K., Lo,K., Offenbacher
              ,J., Danzig,J. and Ducar,M.
TITLE   Creation of genome-wide protein expression libraries using random
              activation of gene expression
JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE   21127151
COMMENT   Contact: Scott J. Cain
              Athersys, Inc.
              3201 Carnegie Ave, Cleveland, OH 44115, USA
              Tel: 216 431 9900
              Fax: 216 361 9596
              Email: scain@athersys.com
              High quality sequence stop: 417.
FEATURES   Location/Qualifiers
     source   1..806
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HT1080"
                     /cell_line="HT1080"
                     /clone_lib="Athersys RAGE Library"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."
BASE COUNT   252 a   187 c   155 g   205 t   7 others
ORIGIN

Query Match   11.9%;   Score 45;   DB 12;   Length 806;
Best Local Similarity 100.0%;   Pred. No. 5.7e-14;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   298 CAGCTATGTTGGTAGCTATAGCAGAAGTCTTCTTGGCAAAGATTC 342
          |||||||||||||||||||||||||||||||||||||||||||||
Db   387 CAGCTATGTTGGTAGCTATAGCAGAAGTCTTCTTGGCAAAGATTC 343

RESULT 5
AQ035618
LOCUS   AQ035618   349 bp   DNA   linear   GSS 11-JUL-1998
DEFINITION   CIT-HSP-2319N3.TF CIT-HSP Homo sapiens genomic clone 2319N3, DNA
              sequence.
ACCESSION   AQ035618
VERSION   AQ035618.1   GI:3301715
KEYWORDS   GSS.
SOURCE   human.
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 349)

AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
              Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
              Simon,M. and Venter,J.C.
TITLE   Use of a random BAC End Sequence Database for Sequence-Ready Map
              Building (1998)
JOURNAL   Unpublished (1998)
COMMENT   Other_GSSs: CIT-HSP-2319N3.TR
              Contact: Mark Adams
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: mdadams@tigr.org
              Clones are available from Research Genetics (info@resgen.com). BAC
              end search page:
              http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
              Seq primer: M13-21
              Class: BAC ends.
FEATURES   Location/Qualifiers
     source   1..349
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2319N3"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT   85 a   69 c   84 g   111 t
ORIGIN

Qy   86 CAAGCTTGAAATGTGTTCTGGAGCTGAAGGATG 119
          |||||||||||||||||||||||||||||||||
Db   31 CAAGCTTGAAATGTGTGTTCTGGAGCTGAAGGATG 64

RESULT 6
AQ035738/c
LOCUS   AQ035738   237 bp   DNA   linear   GSS 11-JUL-1998
DEFINITION   CIT-HSP-2321C21.TF CIT-HSP Homo sapiens genomic clone 2321C21, DNA
              sequence.
ACCESSION   AQ035738
VERSION   AQ035738.1   GI:3301835
KEYWORDS   GSS.
SOURCE   human.
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 237)
AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
              Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
              Simon,M. and Venter,J.C.
TITLE   Use of a random BAC End Sequence Database for Sequence-Ready Map
              Building (1998)
JOURNAL   Unpublished (1998)
COMMENT   Other_GSSs: CIT-HSP-2321C21.TR
              Contact: Mark Adams
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: mdadams@tigr.org
              Clones are available from Research Genetics (info@resgen.com). BAC
              end search page:
              http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
              Seq primer: M13-21
              Class: BAC ends.

Query Match   9.0%;   Score 34;   DB 17;   Length 349;
Best Local Similarity 100.0%;   Pred. No. 7.9e-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
FEATURES             Location/Qualifiers
     source          1..237
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2321C21"
                     /clone_lib="CI1-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT      59 a      47 c      47 g      84 t
ORIGIN

Qy    348 GGGAAGGAAGGTACCATTGGAGAACCATGC 377
          |||||||||||||||||||||||||||||||
Db    224 GGGAAGGAAGGTACCATTGGAGAACCATGC 195
```

```
RESULT 7
LOCUS          BG210731
BG210731/c
DEFINITION     RST30277 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
LOCUS          BG210731          357 bp     mRNA    linear    EST 21-APR-2001
ACCESSION      BG210731
VERSION        BG210731.1    GI:13732418
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 357)
AUTHORS        Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
               Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.
               , J., Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith
               , E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
               , J., Danzig,J. and Ducar,M.
TITLE          Creation of genome-wide protein expression libraries using random
               activation of gene expression
JOURNAL        Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE        21227151
COMMENT        Contact: Scott J. Cain
               Athersys, Inc.
               3201 Carnegie Ave, Cleveland, OH 44115, USA
               Tel: 216 431 9900
               Fax: 216 361 9596
               Email: scain@athersys.com
               High quality sequence stop: 357.
FEATURES             Location/Qualifiers
     source          1..357
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Athersys RAGE Library"
                     /cell_line="HT1080"
                     /note="See 'Creation of Genome-Wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."
Query Match     5.6%;  Score 21;  DB 12;  Length 357;
Best Local Similarity  100.0%;  Pred. No. 1.7;
Matches  21;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    262 GGTGAATGGAGAATGCCAATCCA 282
          ||||||||||||||||||||||
Db     40 GGTGAATGGAGAATGCCAATCCA 20
```

```
RESULT 8
BQ597872/c
LOCUS          BQ597872          440 bp     mRNA    linear    EST 24-JUN-2002
DEFINITION     MI-P-A2-afh-a-04-1-UM_s1 MI-P-A2 Sus scrofa cDNA clone
               MI-P-A2-afh-a-04-1-UM 3', mRNA sequence.
ACCESSION      BQ597872
VERSION        BQ597872.1    GI:21544598
KEYWORDS       EST.
SOURCE         pig.
ORGANISM       Sus scrofa
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE      1 (bases 1 to 440)
AUTHORS        Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE          Normalization and subtraction: two approaches to facilitate gene
               discovery
JOURNAL        Genome Res. 6 (9), 791-806 (1996)
MEDLINE        97044477
COMMENT        Contact: Tuggle CK
               Molecular Genetics Laboratory, Department of Animal Science
               Iowa State University
               201 Kildee Hall, Ames, IA 50011-3150, USA
               Tel: 5152944252
               Fax: 5152942401
               Email: cktuggle@iastate.edu
               Tissue Procurement: Dr. Chris Tuggle, Iowa State University
               cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
               cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
               DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
               Clone Distribution: Researchers may obtain clones from Research
               Genetics (www.resgen.com).
               The following repetitive elements were found in this cDNA
               sequence:1-22, >AT_rich#Low_complexity
               Seq primer: M13 FORWARD
               POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..440
                     /organism="Sus scrofa"
                     /strain="crossbreed"
                     /db_xref="taxon:9823"
                     /clone="MI-P-A2-afh-a-04-1-UM"
                     /clone_lib="MI-P-A2"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-A2
                     library is derived from anterior pituitary at estrus day
                     5. For a detailed description of the library from which
                     this clone was derived, please visit our web site at
                     http://pigest.genome.iastate.edu/.
                     TAG_HOST=MI-P-A2
                     TAG_LIB=MI-P-A2
                     TAG_TISSUE=d_5_anterior_pituitary
                     TAG_SEQ=TCGCGT"
BASE COUNT     127 a      82 c      78 g     153 t
ORIGIN

Qy    350 GAAGGAAGGTACCATTGGAGA 370
          |||||||||||||||||||||
Db    433 GAAGGAAGGTACCATTGGAGA 413
```

```
RESULT 9
BG223756
LOCUS          BG223756          526 bp     mRNA    linear    EST 07-FEB-2001
DEFINITION     IM0018F10a Bovine Mixed Skeletal Muscle cDNA Library Bos taurus
               cDNA 5', mRNA sequence.
ACCESSION      BG223756
VERSION        BG223756.1    GI:12709277
KEYWORDS       EST.
```

Query Match   5.6%; Score 21; DB 14; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match   7.9%; Score 30; DB 17; Length 237;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT   136 a   64 c   49 g   108 t

SOURCE
  ORGANISM Cow.
            Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 526)
  AUTHORS   Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y., Li,G., Murdoch,G.,
            Dixon,W. and Christopherson,B.
  TITLE   cDNA's from bovine mixed skeletal muscle
  JOURNAL   Unpublished (2001)
COMMENT   Contact: Dr. Stephen S. Moore
            Beef Genomics Laboratory
            Dept of AFNS, University of Alberta
            410-Agri/For Centre, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5,
            Canada
            Tel: 780 492 0169
            Fax: 780 492 4265
            Email: smoore@afns.ualberta.ca
            The sequence best matches gb:AF026293 (Homo sapiens chaperonin
            containing t-complex polypeptide 1, beta subunit (Cctb) mRNA,
            complete cds. 12/1998) In GenBank main database at E-value of
            1e-113.
            PCR PRimers
            FORWARD: M13 Forward
            BACKWARD: M13 Reverse
            Seq primer: T3 primer
            High quality sequence stop: 526
            POLYA=Yes.
FEATURES   Location/Qualifiers
     source   1..526
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="Bovine Mixed Skeletal Muscle cDNA library"
                     /sex="two males and one female mixed"
                     /tissue_type="masseter, longissmusclorsi, biceps femoris,
                     senimembrances"
                     /cell_type="muscle"
                     /dev_stage="young adult"
                     /lab_host="Xl1-BlueMRF-strain"
                     /note="Organ: Skeletal muscle; Vector: Uni-ZAPXR; Site_1:
                     EcoR I; Site_2: Xho I"
BASE COUNT   155 a   99 c   119 g   153 t
ORIGIN

Qy   Query Match   5.6%;   Score 21;   DB 12;   Length 526;
          Best Local Similarity   100.0%;   Pred. No. 1.8;
          Matches 21;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   350 GAGGGAAGGTACCATTGGAGA 370
            ||||||||||||||||||||
Db   111 GAGGGAAGGTACCATTGGAGA 131

RESULT 10
AV605455/c
LOCUS   AV605455   541 bp   mRNA   linear   EST 28-NOV-2001
DEFINITION AV605455 Bos taurus kidney fetus Bos taurus cDNA clone E1KI028H04
            3', mRNA sequence.
ACCESSION   AV605455
VERSION   AV605455.1 GI:9735828
KEYWORDS   EST.
SOURCE   cow.
  ORGANISM Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 541)
  AUTHORS   Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
            and Sugimoto,Y.
  TITLE   Establishment of a high throughput EST sequencing system using
            poly(A) tail-removed cDNA libraries and determination of 36,000
            bovine ESTs
  JOURNAL   Nucleic Acids Res. 29 (22), E108 (2001)

MEDLINE   21570554
COMMENT   Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shirakawa Institute of Animal Genetics
            Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@cocoa.ocn.ne.jp
            Single pass sequencing.
            This clone was obtained from a polyA-deleted cDNA library.
FEATURES   Location/Qualifiers
     source   1..541
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone="E1KI028H04"
                     /clone_lib="Bos taurus kidney fetus"
                     /tissue_type="kidney"
                     /dev_stage="fetus"
                     /lab_host="DH10B"
                     /note="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
                     was deleted from a NotI site"
BASE COUNT   139 a   136 c   110 g   153 t   3 others
ORIGIN

Qy   Query Match   5.6%;   Score 21;   DB 10;   Length 541;
          Best Local Similarity   100.0%;   Pred. No. 1.8;
          Matches 21;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   350 GAGGGAAGGTACCATTGGAGA 370
            |||||||||||||||||||||
Db   331 GAGGGAAGGTACCATTGGAGA 311

RESULT 11
AV605117/c
LOCUS   AV605117   545 bp   mRNA   linear   EST 28-NOV-2001
DEFINITION AV605117 Bos taurus kidney fetus Bos taurus cDNA clone E1KI026H04
            3', mRNA sequence.
ACCESSION   AV605117
VERSION   AV605117.1 GI:9735490
KEYWORDS   EST.
SOURCE   cow.
  ORGANISM Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 545)
  AUTHORS   Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso.H.
            and Sugimoto,Y.
  TITLE   Establishment of a high throughput EST sequencing system using
            poly(A) tail-removed cDNA libraries and determination of 36,000
            bovine ESTs
  JOURNAL   Nucleic Acids Res. 29 (22), E108 (2001)
  MEDLINE   21570554
COMMENT   Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shirakawa Institute of Animal Genetics
            Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@cocoa.ocn.ne.jp
            Single pass sequencing.
            This clone was obtained from a polyA-deleted cDNA library.
FEATURES   Location/Qualifiers
     source   1..545
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone="E1KI026H04"
                     /clone_lib="Bos taurus kidney fetus"
                     /tissue_type="kidney"
                     /dev_stage="fetus"
                     /lab_host="DH10B"
                     /note="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A

```
BASE COUNT        136 a    145 c    110 g    154 t
ORIGIN
                  was deleted from a NotI site"

Qy    350 GAAGGAAGGTACCATTGGAGA 370
          |||||||||||||||||||||
Db    289 GAAGGAAGGTACCATTGGAGA 269

Query Match         5.6%; Score 21; DB 10; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 12
LOCUS       BF713089           547 bp    mRNA    linear    EST 02-JAN-2001
BF713089/c
DEFINITION  MI-P-H3-adm-h-03-1-UM.s1 MI-P-H3 Sus scrofa cDNA clone
            MI-P-H3-adm-h-03-1-UM 3', mRNA sequence.
ACCESSION   BF713089
VERSION     BF713089.1  GI:12012570
KEYWORDS    EST.
SOURCE      pig.
ORGANISM    Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1 (bases 1 to 547)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
COMMENT     Contact: Tuggle CK
            Molecular Genetics Laboratory, Department of Animal Science
            Iowa State University
            201 Kildee Hall, Ames, IA 50011-3150, USA
            Tel: 515294252
            Fax: 515942401
            Email: cktuggle@iastate.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            non-normalized hypothalamus at estrus day 12 library cDNA Library
            Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science
            Research Center, Department of Animal Science, University of
            Missouri-Columbia, 65211 Clone distribution: clones will be
            available through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES    Location/Qualifiers
   source           1..547
                    /organism="Sus scrofa"
                    /strain="crossbreed"
                    /db_xref="taxon:9823"
                    /clone="MI-P-H3-adm-h-03-1-UM"
                    /clone_lib="MI-P-H3"
                    /lab_host="DH10B (Life Technologies)"
                    /note=Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-H3
                    library is derived from hypothalamus at estrus day 12.
                    For a detailed description of the library from which this
                    clone was derived, please visit our web site at
                    http://pigest.genome.iastate.edu/.
                    TAG_LIB=MI-P-H3
                    TAG_TISSUE=hypothalamus at estrus day 12
                    TAG_SEQ=GGTTAA"
BASE COUNT        157 a    107 c    102 g    181 t
ORIGIN
```

```
Qy    350 GAAGGAAGGTACCATTGGAGA 370
          |||||||||||||||||||||
Db    450 GAAGGAAGGTACCATTGGAGA 430

Query Match         5.6%; Score 21; DB 10; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 13
LOCUS       AW464207           569 bp    mRNA    linear    EST 24-FEB-2000
DEFINITION  BP230015A10F9 Soares normalized bovine placenta Bos taurus cDNA
            clone BP230015A10F9 5', mRNA sequence.
ACCESSION   AW464207
VERSION     AW464207.1  GI:7034375
KEYWORDS    EST.
SOURCE      cow.
ORGANISM    Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 569)
AUTHORS     Lewin,H.A., Soares,M.B., Rebelz,M., Pardinas,J., Liu,L. and Larson
            ,J.H.
TITLE       Bovine ESTs
JOURNAL     Unpublished (2000)
COMMENT     Contact: Lewin, H. A.
            W. M. Keck Center for Comparative and Functional Genomics
            University of Illinois at Urbana-Champaign
            340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
            61801, USA
            Tel: 217 333 5998
            Fax: 217 244 5617
            Email: h-lewin@uiuc.edu
            Funding for cattle EST sequencing was provided by the USDA National
            Research Initiative, Animal Genome Research Grant AG 99-3205-8534
            to R. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
            from Washington University Genome Center. Vector Trimml.g:
            Cross_match from Washington University Genome Center PHRAP suite.
            Sequences submitted are vector free and at least 200 bp in length.
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Insert Length: 569    Std Error: 0.00
            Plate: BP230015A10    row: F column: 9
            Seq Primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 569.
FEATURES    Location/Qualifiers
   source           1..569
                    /organism="Bos taurus"
                    /db_xref="taxon:9913"
                    /sex="female"
                    /lab_host="DH10B"
                    /clone_lib="Soares normalized bovine placenta"
                    /clone="BP230015A10F9"
                    /note="Organ: placenta; Vector: pT7T3Pac; Site_1: EcoRI;
                    Site_2: NotI; The cDNA library was constructed and
                    Soares laboratory and it was contributed by the
                    as described by Bonaldo, M.F., Lennon, G. and Soares,
                    M.B. (1996), Genome Research 6(9): 791-806."
BASE COUNT        160 a    112 c    144 g    150 t    3 others
ORIGIN

Qy    350 GAAGGAAGGTACCATTGGAGA 370
          |||||||||||||||||||||
Db    232 GAAGGAAGGTACCATTGGAGA 252

Query Match         5.6%; Score 21; DB 10; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 14
LOCUS       AZ364217           589 bp    DNA    linear    GSS 02-OCT-2000
AZ364217
```

DEFINITION IM0110P18F Mouse 10kb plasmid UUGC1M0110P18 F, DNA sequence.
ACCESSION   A2364217
VERSION   A2364217.1 GI:10477917
KEYWORDS   GSS.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 589)
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D. Weiss,R.
TITLE   Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
            University of Utah
            University of Utah Genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0110 row: P column: 18
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 589.
FEATURES   Location/Qualifiers
     source   1..589
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0110P18"
                     /clone_lib="Mouse 10kb plasmid library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, Ti-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|14732114|gb|AF129072.1|), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT   130 a   141 c   110 g   208 t
ORIGIN

RESULT 15
LOCUS   BQ600629   611 bp   mRNA   linear   EST 24-JUN-2002

Query Match   5.6%;   Score 21;   DB 17;   Length 589;
Best Local Similarity   100.0%;   Pred. No. 1.8;
Matches 21;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   219 CTTTGACTGTTGAAAAGAGAC 239
          |||||||||||||||||||||
Db   524 CTTTGACTGTTGAAAAGAGAC 544

DEFINITION MI-P-E7-agz-d-10-1-UM_s1 MI-P-E7 Sus scrofa cDNA clone
            MI-P-E7-agz-d-10-1-UM 3', mRNA sequence.
ACCESSION   BQ600629
VERSION   BQ600629.1 GI:21547355
KEYWORDS   EST.
SOURCE   pig.
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1 (bases 1 to 611)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE   Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
COMMENT   Contact: Tuggle CK
            Molecular Genetics Laboratory, Department of Animal Science
            Iowa State University
            201 Kildee Hall, Ames, IA 50011-3150, USA
            Tel: 5152944252
            Fax: 5152942401
            Email: cktuggle@lastate.edu
            Tissue Procurement: Dr. Chris Tuggle, Iowa State University
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            The following repetitive elements were found in this cDNA
            sequence: 1-22, >AT_rich#Low_complexity
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES   Location/Qualifiers
     source   1..611
                     /organism="Sus scrofa"
                     /strain="crossbreed"
                     /db_xref="taxon:9823"
                     /clone="MI-P-E7-agz-d-10-1-UM"
                     /clone_lib="MI-P-E7"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pSPORT1; Site_1: Not I; Site_2: EcoRI; The
                     MI-P-E7 library is derived from the following tissue(s):
                     preelongation day 12 conceptus. For a detailed description
                     of the library from which this clone was derived, please
                     visit our web site at http://pigest.genome.lastate.edu/.
                     TAG_LIB=MI-P-E7
                     TAG_LIB=MI-P-E7
                     TAG_TISSUE=preelongation_d_12_conceptus
                     TAG_SEQ=GTGAGA."
BASE COUNT   164 a   123 c   118 g   206 t
ORIGIN

Query Match   5.6%;   Score 21;   DB 14;   Length 611;
Best Local Similarity   100.0%;   Pred. No. 1.8;
Matches 21;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   350 GAAGGAAGGTACCATTGGAGA 370
          |||||||||||||||||||||
Db   434 GAAGGAAGGTACCATTGGAGA 414

Search completed: January 31, 2003, 05:14:00
Job time : 2258 secs

```
OM nucleic - nucleic search, using sw model

Run on:            January 31, 2003, 01:17:34 ; Search time 267 Seconds
                                               (without alignments)
                                               3188.222 Million cell updates/sec

Title:             US-09-803-719-222
Perfect score:     378
Sequence:          1 tacgggctgcttaagacgact............taccattggagaacatgcg 378

Scoring table:     OLIGO_NUC
                   Gapop 60.0 , Gapext 60.0

Searched:          2185239 seqs, 112599159 residues

Word size :        0

Total number of hits satisfying chosen parameters:    4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :  N_Geneseq_101002:*
    1:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
    2:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
    3:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
    4:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
    5:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
    6:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
    7:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
    8:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
    9:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
   10:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
   11:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
   12:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
   13:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
   14:  /SIDS2/gcgdata/geneseq/geneseqn-embl1/NA1993.DAT:*
   15:  /SIDS2/gcgdata/geneseq/geneseqn-embl1/NA1994.DAT:*
   16:  /SIDS2/gcgdata/geneseq/geneseqn-embl1/NA1995.DAT:*
   17:  /SIDS2/gcgdata/geneseq/geneseqn-embl1/NA1996.DAT:*
   18:  /SIDS2/gcgdata/geneseq/geneseqn-embl1/NA1997.DAT:*
   19:  /SIDS2/gcgdata/geneseq/geneseqn-embl1/NA1998.DAT:*
   20:  /SIDS2/gcgdata/geneseq/geneseqn-embl1/NA1999.DAT:*
   21:  /SIDS2/gcgdata/geneseq/geneseqn-embl1/NA2000.DAT:*
   22:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
   23:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
   24:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	378	100.0	378	22	AAS37164	Novel human diagno
2	238	63.0	322	22	AAS37237	Novel human diagno
3	238	63.0	322	22	AAS37218	Novel human diagno
4	232	61.4	318	22	AAS37242	Novel human diagno
5	19	5.0	411	22	ABA20368	Human nervous syst
6	19	5.0	527	22	ABA61650	Human foetal liver
7	19	5.0	527	22	ABA29309	Probe #7775 for ge
8	19	5.0	527	22	AAK09951	Human brain expres
9	19	5.0	527	22	AAK35845	Human bone marrow
10	19	5.0	527	22	AAI17155	Probe #7088 for ge
11	19	5.0	527	22	AAI41562	Probe #10248 used
12	19	5.0	527	24	ABS10078	Human genome-deriv
13	19	5.0	9929	22	ABA17906	Human nervous syst
14	18	4.8	249	22	AAK58913	Human immune/haema
15	18	4.8	333	22	AAK74060	Human immune/haema
16	18	4.8	721	24	AAS62111	Porcine muscular s
17	18	4.8	721	24	AAS62084	Porcine muscular s
18	18	4.8	5090	23	ABL26677	Drosophila melanog
19	18	4.8	1768	23	ABL20676	Drosophila melanog
20	18	4.8	8638	22	AAK68372	Porcine muscular s
21	18	4.8	26555	22	AAK68605	Human immune/haema
22	18	4.8	26555	22	AAI62833	Human immune/haema
23	18	4.8	26555	22	AAC18996	Human genomic DNA
24	17	4.5	281	21	AAK54310	Human secreted pro
25	17	4.5	309	22	ABV51017	Human haematologic
26	17	4.5	597	23	AAV88365	Human prostate exp
27	17	4.5	627	20	ABQ90901	EST clone DD211.
28	17	4.5	738	24	AAI96650	M. capsulatus gene
29	17	4.5	740	22	AAC43174	Human neuroblastom
30	17	4.5	783	21	AAI99451	Arabidopsis thalia
31	17	4.5	999	22	AAH51589	Human polynucleoti
32	17	4.5	1000	21	AAH51435	Human UGT1A7/relat
33	17	4.5	1001	21	AAH51436	Human UGT1A7 relat
34	17	4.5	1001	21	AAH51455	Human UGT1A7 relat
35	17	4.5	1001	21	ABA69091	Human foetal liver
36	17	4.5	1041	22	AAK43200	Human bone marrow
37	17	4.5	1041	22	AAI49277	Human bone marrow
38	17	4.5	1041	22	AAI09567	Probe #17963 used
39	17	4.5	1041	24	ABS17281	Probe #9558 used t
40	17	4.5	1080	24	ABN91801	Human genome-deriv
41	17	4.5	1180	23	ABL17933	Drosophila melanog
42	17	4.5	1290	21	AAC46337	Staphylococcus epi
43	17	4.5	1290	21	AAC48898	Arabidopsis thalia
44	17	4.5	1293	21	ABA69091	Arabidopsis thalia
45	17	4.5	1294	21	AAC42098	Arabidopsis thalia

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

```
RESULT 1
AAS37164
ID   AAS37164 standard; cDNA; 378 BP.
XX
AC   AAS37164;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Novel human diagnostic and therapeutic gene #222.
XX
KW   Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200166753-A2.
XX
PD   13-SEP-2001.
XX
PF   09-MAR-2001; 2001WO-US07787.
XX
PR   09-MAR-2000; 2000US-0188609.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI   Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI   Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI   Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
DR   WPI; 2001-530177/58.
```

PT   New polynucleotides and polypeptides, useful for diagnosis and
PT   treatment of breast, lung and colon cancer -
XX
PS   Claim 1; Page 652; 1193pp; English.
XX
CC   The invention relates to new polynucleotides and polypeptides, useful for
CC   diagnosis and treatment of breast, lung and colon cancer. The sequences
CC   can be used in detecting differentially expressed genes correlated with a
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample derived from a
CC   cell suspected of being cancerous. They can also be used to inhibit
CC   tumour growth by modulating expression of a gene product. AAS36943-
CC   AAS39938 represent novel human diagnostic and therapeutic coding
CC   sequences of the invention.
XX
SQ   Sequence 378 BP; 100 A; 76 C; 101 G; 101 T; 0 other;

Query Match   100.0%;   Score 378;   DB 22;   Length 378;
Best Local Similarity   100.0%;   Pred. No. 1.6e-187;
Matches 378;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy      1 TACGGCTGCTTAAGACGACTTAAGGGGGGAATGACCGAGCCGGCTCTTAGAGAGAACATATGG   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TACGGCTGCTTAAGACGACTTAAGGGGGGAATGACCGAGCCGGCTCTTAGAGAGAACATATGG   60

Qy     61 AAAACACCCAAGCCGGAGTCTCTCACAAGCTTGAAGTGTGTGTTCTGGGAGCTGAAGGATGC  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAAACACCCAAGCCGGAGTCTCTCACAAGCTTGAAGTGTGTGTTCTGGGAGCTGAAGGATGC  120

Qy    121 ACGGTTGTTAAGCCCCTGTTCTTTTCCGTTGTTGTTTAATCTAATAGTTCTTTGGAATAAAC  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACGGTTGTTAAGCCCCTGTTCTTTTCCGTTGTTGTTTAATCTAATAGTTCTTTGGAATAAAC  180

Qy    181 CTCCCTGCCAAGTACTACTTGGTTTTATGCTCACAATGCTTTGACTGTTGAAAAGAGACC  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTCCCTGCCAAGTACTACTTGGTTTTATGCTCACAATGCTTTGACTGTTGAAAAGAGACC  240

Qy    241 TTTGGCACACATTGAAGGGGATGGTGATGGAGATGCCAATCCATGGAAATCAGGTGGCGCAG  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTTGGCACACATTGAAGGGGATGGTGATGGAGATGCCAATCCATGGAAATCAGGTGGCGCAG  300

Qy    301 CTATGTTGGTAGCTATAGCAAGAAGTCTTCTTGGCAAAGATTCCTCCCGGGAAGGAAGGTA  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTATGTTGGTAGCTATAGCAAGAAGTCTTCTTGGCAAAGATTCCTCCCGGGAAGGAAGGTA  360

Qy    361 CCATTGGAGAACCATGCG  378
          ||||||||||||||||||
Db    361 CCATTGGAGAACCATGCG  378
```

RESULT 2
AAS37237
ID   AAS37237 standard; cDNA; 322 BP.
XX
AC   AAS37237;
XX
DT   17-DEC-2001   (first entry)
XX
DE   Novel human diagnostic and therapeutic gene #295.
XX
KW   Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200166753-A2.
XX
PD   13-SEP-2001.
XX
PF   09-MAR-2001; 2001WO-US07787.
XX
PR   09-MAR-2000; 2000US-0188609.
XX
PA   (CHIR) CHIRON CORP.

PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,   Escobedo J,   Innis MA,   Garcia PD,   Sudduth-Klinger J;
PI   Reinhard C,   Randazzo F,   Kennedy GC,   Pot D,   Lamson G;
PI   Drmanac R,   Crkvenjakov R,   Dickson M,   Drmanac S,   Labat I;
PI   Leshkowitz D,   Kita D,   Garcia V,   Jones WL,   Stache-Crain B;
XX
DR   WPI; 2001-530177/58.
XX
CC   The invention relates to new polynucleotides and polypeptides, useful for
CC   diagnosis and treatment of breast, lung and colon cancer. The sequences
CC   can be used in detecting differentially expressed genes correlated with a
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample derived from a
CC   cell suspected of being cancerous. They can also be used to inhibit
CC   tumour growth by modulating expression of a gene product. AAS36943-
CC   AAS39938 represent novel human diagnostic and therapeutic coding
CC   sequences of the invention.
XX
PS   Claim 1; Page 670; 1193pp; English.
XX
PT   New polynucleotides and polypeptides, useful for diagnosis and
PT   treatment of breast, lung and colon cancer -
XX
ID   New polynucleotides and polypeptides, useful for diagnosis and
XX
SQ   Sequence 322 BP; 90 A; 64 C; 78 G; 89 T; 1 other;

Query Match   63.0%;   Score 238;   DB 22;   Length 322;
Best Local Similarity   100.0%;   Pred. No. 2.6e-114;
Matches 238;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy     59 GGAAAACACCAAGCCGAGCCGGAGTCTCTCACAAGCTTGAATGTGTGTGTTCTGGAGCTGAAGGAT  118
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     67 GGAAAACACCAAGCCGGAGTCTCTCACAAGCTTGAATGTGTGTGTTCTGGACCTGAAGGAT  126

Qy    119 GCACGGTTGTTAAGCCCCTGTTCTTCTTTCCGTTGTTGTTTAATCTAATGTTCTTTGGAATAAA  178
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    127 GCACGGTTGTTAAGCCCCTGTTCTTCTTTCCGTTGTTGTTTAATCTAATGTTCTTTGGAATAAA  186

Qy    179 ACCTCCCTGCCAAGTAGTACTTGGTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGA  238
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187 ACCTCCCTGCCAAGTAGTACTTGGTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGA  246

Qy    239 CCTTTGGCACACATTGAAAGGGATGGTGATGGAGATGCCAAATCCATGGAATCAGGTGCGC  296
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    247 CCTTTGCGCACACATTGAAAGGGATGGTGATGGAGATGCCAAATCCATGGAATCAGGTGCGC  304
```

RESULT 3
AAS37218
ID   AAS37218 standard; cDNA; 327 BP.
XX
AC   AAS37218;
XX
DT   17-DEC-2001   (first entry)
XX
DE   Novel human diagnostic and therapeutic gene #276.
XX
KW   Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200166753-A2.
XX
PD   13-SEP-2001.
XX
PF   09-MAR-2001; 2001WO-US07787.
XX
PR   09-MAR-2000; 2000US-0188609.
XX
PA   (CHIR) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,   Escobedo J,   Innis MA,   Garcia PD,   Sudduth-Klinger J;

RESULT 4

ID   AAS37242

AC   AAS37242;

DT   17-DEC-2001 (first entry)

DE   Novel human diagnostic and therapeutic gene #300.

KW   Human; cancer; breast; lung; colon; prostate; diagnostic; ss.

OS   Homo sapiens.

PN   WO200166753-A2.

PD   13-SEP-2001.

PF   09-MAR-2001; 2001WO-US007787.

PR   09-MAR-2000; 2000US-0188609.

PA   (CHIR) CHIRON CORP.

PA   (HYSE-) HYSEQ INC.

PI   Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI   Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI   Drmanac R, Crkvenjakov R, Dickson M, Dmanac S, Labat I;
PI   Leshkovitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;

XX

DR   WPI; 2001-530177/58.

XX

PT   New polynucleotides and polypeptides, useful for diagnosis and
PT   treatment of breast, lung and colon cancer -

XX

PS   Claim 1; Page 665; 1193pp; English.

XX

CC   The invention relates to new polynucleotides and polypeptides, useful for
CC   diagnosis and treatment of breast, lung and colon cancer. The sequences
CC   can be used in detecting differentially expressed genes correlated with a
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample derived from a
CC   cell suspected of being cancerous. They can also be used to inhibit
CC   tumour growth by modulating expression of a gene product. AAS36943-
CC   AAS39338 represent novel human diagnostic and therapeutic coding
CC   sequences of the invention.

XX

SQ   Sequence 327 BP; 92 A; 64 C; 81 G; 90 T; 0 other;

Query Match   63.0%; Score 238; DB 22; Length 327;
Best Local Similarity   100.0%;
Matches 238; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   59 GGAAAACACCCAAGCCGGAGTCTCTCACAAGCTTGAATGTGTGTTCTGGAGCTGAAGGAT 118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   52 GGAAAACACCCAAGCCGGAGTCTCTCACAAGCTTGAATGTGTGTTCTGGAGCTGAAGGAT 111

Qy   119 GCACGGTTGTTAAGCCCCCTGTTCTTTCCGTTGTTGTTTAATCTAANTGTTCTTTGGAATAAAA 178
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   112 GCACGGTTGTTAAGCCCCCTGTTCTTTCCGTTGTTGTTTAATCTAANTGTTCTTTGGAATAAAA 171

Qy   179 ACCTCCCTGCCAAGTAGTACTTGGTTTTTATGCTTCAACATGCTTTGACTGTTGAAAAGAGA 238
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   172 ACCTCCCTGCCAAGTAGTACTTGGTTTTTATGCTTCAACATGCTTTGACTGTTGAAAAGAGA 231

Qy   239 CCTTTGGCACACATTGAAAGGGATGGTGATGGAGAGATGCCAATCCATGGAATCAGGTGGC 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   232 CCTTTGGCACACATTGAAAGGGATGGTGATGGAGAGATGCCAATCCATGGAATCAGGTGGC 289

RESULT 5

ID   ABA20368

AC   ABA20368;

DT   23-JAN-2002 (first entry)

DE   Human nervous system related polynucleotide SEQ ID NO 12699.

KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cereboprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

OS   Homo sapiens.

PN   WO200159063-A2.

PD   16-AUG-2001.

PF   17-JAN-2001; 2001WO-US01334.

PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0185350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.

XX

DR   WPI; 2001-530177/58.

XX

PT   New polynucleotides and polypeptides, useful for diagnosis and
PT   treatment of breast, lung and colon cancer -

XX

PS   Claim 1; Page 671; 1193pp; English.

XX

CC   The invention relates to new polynucleotides and polypeptides, useful for
CC   diagnosis and treatment of breast, lung and colon cancer. The sequences
CC   can be used in detecting differentially expressed genes correlated with a
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample derived from a
CC   cell suspected of being cancerous. They can also be used to inhibit
CC   tumour growth by modulating expression of a gene product. AAS36943-
CC   AAS39338 represent novel human diagnostic and therapeutic coding
CC   sequences of the invention.

XX

SQ   Sequence 318 BP; 90 A; 63 C; 76 G; 89 T; 0 other;

Query Match   61.8%; Score 232; DB 22; Length 318;
Best Local Similarity   100.0%;
Matches 232; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   59 GGAAAACACCCAAGCCGGAGTCTCTCACAAGCTTGAATGTGTGTTCTGGAGCTGAAGGAT 118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   67 GGAAAACACCCAAGCCGGAGTCTCTCACAAGCTTGAATGTGTGTTCTGGAGCTGAAGGAT 126

Qy   119 GCACGGTTGTTAAGCCCCCTGTTCTTTCCGTTGTTGTTTAATCTAANTGTTCTTTGGAATAAAA 178
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   127 GCACGGTTGTTAAGCCCCCTGTTCTTTCCGTTGTTGTTTAATCTAANTGTTCTTTGGAATAAAA 186

Qy   179 ACCTCCCTGCCAAGTAGTACTTGGTTTTTATGCTTCAACATGCTTTGACTGTTGAAAAGAGA 238
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   187 ACCTCCCTGCCAAGTAGTACTTGGTTTTTATGCTTCAACATGCTTTGACTGTTGAAAAGAGA 246

Qy   239 CCTTTGGCACACATTGAAAGGGATGGTGATGGAGAGATGCCAATCCATGGAATCA 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   247 CCTTTGGCACACATTGAAAGGGATGGTGATGGAGAGATGCCAATCCATGGAATCA 298

```
PR   18-APR-2000;   2000US-0198123.
PR   19-MAY-2000;   2000US-0205515.
PR   07-JUN-2000;   2000US-0209467.
PR   28-JUN-2000;   2000US-0214886.
PR   30-JUN-2000;   2000US-0215135.
PR   07-JUL-2000;   2000US-0216647.
PR   07-JUL-2000;   2000US-0216880.
PR   11-JUL-2000;   2000US-0217487.
PR   11-JUL-2000;   2000US-0217496.
PR   14-JUL-2000;   2000US-0218290.
PR   26-JUL-2000;   2000US-0220963.
PR   26-JUL-2000;   2000US-0220964.
PR   14-AUG-2000;   2000US-0224518.
PR   14-AUG-2000;   2000US-0224519.
PR   14-AUG-2000;   2000US-0225213.
PR   14-AUG-2000;   2000US-0225214.
PR   14-AUG-2000;   2000US-0225266.
PR   14-AUG-2000;   2000US-0225267.
PR   14-AUG-2000;   2000US-0225268.
PR   14-AUG-2000;   2000US-0225270.
PR   14-AUG-2000;   2000US-0225447.
PR   14-AUG-2000;   2000US-0225757.
PR   14-AUG-2000;   2000US-0225758.
PR   18-AUG-2000;   2000US-0225759.
PR   22-AUG-2000;   2000US-0226279.
PR   22-AUG-2000;   2000US-0226681.
PR   22-AUG-2000;   2000US-0226868.
PR   23-AUG-2000;   2000US-0227182.
PR   30-AUG-2000;   2000US-0227009.
PR   01-SEP-2000;   2000US-0228924.
PR   01-SEP-2000;   2000US-0229287.
PR   01-SEP-2000;   2000US-0229343.
PR   01-SEP-2000;   2000US-0229344.
PR   01-SEP-2000;   2000US-0229345.
PR   05-SEP-2000;   2000US-0229509.
PR   05-SEP-2000;   2000US-0229513.
PR   06-SEP-2000;   2000US-0230437.
PR   06-SEP-2000;   2000US-0230438.
PR   08-SEP-2000;   2000US-0231242.
PR   08-SEP-2000;   2000US-0231243.
PR   08-SEP-2000;   2000US-0231244.
PR   08-SEP-2000;   2000US-0231413.
PR   08-SEP-2000;   2000US-0231414.
PR   08-SEP-2000;   2000US-0232080.
PR   08-SEP-2000;   2000US-0232081.
PR   12-SEP-2000;   2000US-0231968.
PR   14-SEP-2000;   2000US-0232397.
PR   14-SEP-2000;   2000US-0232398.
PR   14-SEP-2000;   2000US-0232399.
PR   14-SEP-2000;   2000US-0232400.
PR   14-SEP-2000;   2000US-0232401.
PR   14-SEP-2000;   2000US-0233063.
PR   14-SEP-2000;   2000US-0233064.
PR   14-SEP-2000;   2000US-0233065.
PR   21-SEP-2000;   2000US-0234223.
PR   21-SEP-2000;   2000US-0234274.
PR   25-SEP-2000;   2000US-0234997.
PR   25-SEP-2000;   2000US-0234998.
PR   26-SEP-2000;   2000US-0235484.
PR   27-SEP-2000;   2000US-0235834.
PR   27-SEP-2000;   2000US-0235836.
PR   29-SEP-2000;   2000US-0236327.
PR   29-SEP-2000;   2000US-0236367.
PR   29-SEP-2000;   2000US-0236368.
PR   29-SEP-2000;   2000US-0236369.
PR   29-SEP-2000;   2000US-0236370.
PR   02-OCT-2000;   2000US-0236802.
PR   02-OCT-2000;   2000US-0237037.
PR   02-OCT-2000;   2000US-0237038.
PR   02-OCT-2000;   2000US-0237039.
PR   02-OCT-2000;   2000US-0237040.
PR   13-OCT-2000;   2000US-0239935.
PR   13-OCT-2000;   2000US-0239937.

PR   20-OCT-2000;   2000US-0240960.
PR   20-OCT-2000;   2000US-0241785.
PR   20-OCT-2000;   2000US-0241786.
PR   20-OCT-2000;   2000US-0241787.
PR   20-OCT-2000;   2000US-0241808.
PR   20-OCT-2000;   2000US-0241809.
PR   20-OCT-2000;   2000US-0241826.
PR   20-OCT-2000;   2000US-0242221.
PR   01-NOV-2000;   2000US-0244617.
PR   08-NOV-2000;   2000US-0244474.
PR   08-NOV-2000;   2000US-0244475.
PR   08-NOV-2000;   2000US-0244476.
PR   08-NOV-2000;   2000US-0244477.
PR   08-NOV-2000;   2000US-0244478.
PR   08-NOV-2000;   2000US-0245523.
PR   08-NOV-2000;   2000US-0245524.
PR   08-NOV-2000;   2000US-0245525.
PR   08-NOV-2000;   2000US-0245526.
PR   08-NOV-2000;   2000US-0245527.
PR   08-NOV-2000;   2000US-0245528.
PR   08-NOV-2000;   2000US-0245532.
PR   08-NOV-2000;   2000US-0246609.
PR   08-NOV-2000;   2000US-0246610.
PR   08-NOV-2000;   2000US-0246611.
PR   08-NOV-2000;   2000US-0246613.
PR   17-NOV-2000;   2000US-0249207.
PR   17-NOV-2000;   2000US-0249208.
PR   17-NOV-2000;   2000US-0249209.
PR   17-NOV-2000;   2000US-0249210.
PR   17-NOV-2000;   2000US-0249211.
PR   17-NOV-2000;   2000US-0249212.
PR   17-NOV-2000;   2000US-0249213.
PR   17-NOV-2000;   2000US-0249214.
PR   17-NOV-2000;   2000US-0249215.
PR   17-NOV-2000;   2000US-0249216.
PR   17-NOV-2000;   2000US-0249217.
PR   17-NOV-2000;   2000US-0249218.
PR   17-NOV-2000;   2000US-0249244.
PR   17-NOV-2000;   2000US-0249245.
PR   17-NOV-2000;   2000US-0249264.
PR   17-NOV-2000;   2000US-0249265.
PR   17-NOV-2000;   2000US-0249297.
PR   17-NOV-2000;   2000US-0249299.
PR   17-NOV-2000;   2000US-0249300.
PR   01-DEC-2000;   2000US-0249391.
PR   01-DEC-2000;   2000US-0250391.
PR   01-DEC-2000;   2000US-0251030.
PR   05-DEC-2000;   2000US-0251160.
PR   05-DEC-2000;   2000US-0251988.
PR   05-DEC-2000;   2000US-0256719.
PR   06-DEC-2000;   2000US-0251479.
PR   08-DEC-2000;   2000US-0251856.
PR   08-DEC-2000;   2000US-0251868.
PR   08-DEC-2000;   2000US-0251669.
PR   08-DEC-2000;   2000US-0251989.
PR   08-DEC-2000;   2000US-0251990.
PR   11-DEC-2000;   2000US-0254097.
PR   05-JAN-2001;   2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Barash SC,   Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
     useful for preventing, diagnosing and/or treating nervous system
     cancers and metastases -
PS   Disclosure; SEQ ID NO 12699; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
     (ABB14678-ABB18001) useful for preventing, treating or ameliorating
     medical conditions e.g. by protein or gene therapy. The genes are
```

Left column

```
CC     isolated from a range of human tissues disclosed in the specification.
CC     the nucleic acids, proteins, antibodies and (ant)agonists are useful
CC     in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC     and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC     marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC     (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC     haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC     disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC     colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC     (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC     epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC     and parasitic infections.
CC     Note: The sequence data for this patent did not form part of the
CC     printed specification, but was obtained in electronic format directly
CC     from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ     Sequence 411 BP; 105 A; 79 C; 82 G; 145 T; 0 other;

Query Match     5.0%; Score 19; DB 22; Length 411;
Best Local Similarity 100.0%; Pred. No. 9;
Matches    19; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

Qy       101 GTTCTGGAGCTGAAGGATG 119
             |||||||||||||||||||
Db       267 GTTCTGGAGCTGAAGGATG 285


RESULT 6
ABA61650
ID     ABA61650 standard; DNA; 527 BP.
XX
AC     ABA61650;
XX
DT     01-FEB-2002  (first entry)
XX
DE     Human foetal liver single exon nucleic acid probe #9955.
XX
KW     Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS     Homo sapiens.
XX
PN     WO200157277-A2.
XX
PD     09-AUG-2001.
XX
PF     30-JAN-2001; 2001WO-US00669.
XX
PR     04-FEB-2000; 2000US-0180312.
PR     26-MAY-2000; 2000US-0207456.
PR     30-JUN-2000; 2000US-0608408.
PR     03-AUG-2000; 2000US-0632366.
PR     21-SEP-2000; 2000US-0234687.
PR     27-SEP-2000; 2000US-0236359.
PR     04-OCT-2000; 2000GB-0024263.
XX
PA     (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI     Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR     WPI; 2001-483447/52.
XX
PT     Human genome-derived single exon nucleic acid probes useful for
PT     analyzing gene expression in human fetal liver -
XX
PS     Claim 1; SEQ ID NO 9955; 639pp + sequence listing; English.
XX
CC     The invention relates to a single exon nucleic acid probe for
CC     measuring human gene expression in a sample derived from human foetal
CC     liver. The single exon nucleic acid probes may be used for predicting,
CC     measuring and displaying gene expression in samples derived from human
CC     fetal liver. The present sequence is a single exon nucleic acid
CC     probe of the invention.
CC     Note: The sequence data for this patent did not form part of the
```

Right column

```
CC     printed specification, but was obtained in electronic format directly
CC     from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ     Sequence 527 BP; 175 A; 99 C; 110 G; 143 T; 0 other;

Query Match     5.0%; Score 19; DB 22; Length 527;
Best Local Similarity 100.0%; Pred. No. 9;
Matches    19; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

Qy       162 TGTTCTTTGGAATAAAAAC 180
             |||||||||||||||||||
Db       477 TGTTCTTTGGAATAAAAAC 495


RESULT 7
ABA29309
ID     ABA29309 standard; DNA; 527 BP.
XX
AC     ABA29309;
XX
DT     23-JAN-2002  (first entry)
XX
DE     Probe #7775 for gene expression analysis in human heart cell sample.
XX
KW     Human; gene expression; heart; microarray; vascular system; probe;
KW     cardiovascular disease; hypertension; cardiac arrhythmia;
KW     congenital heart disease; ss.
XX
OS     Homo sapiens.
XX
PN     WO200157274-A2.
XX
PD     09-AUG-2001.
XX
PF     30-JAN-2001; 2001WO-US00666.
XX
PR     04-FEB-2000; 2000US-0180312.
PR     26-MAY-2000; 2000US-0207456.
PR     30-JUN-2000; 2000US-0608408.
PR     03-AUG-2000; 2000US-0632366.
PR     21-SEP-2000; 2000US-0234687.
PR     27-SEP-2000; 2000US-0236359.
PR     04-OCT-2000; 2000GB-0024263.
XX
PA     (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI     Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR     WPI; 2001-488899/53.
XX
PT     Single exon nucleic acid probes for analyzing gene expression in human
PT     hearts -
XX
PS     Claim 1; SEQ ID NO 7775; 530pp; English.
XX
CC     The present invention relates to single exon nucleic acid probes for
CC     measuring human gene expression in a sample derived from human heart. The
CC     present sequence is one such probe. The probes may be used for
CC     predicting, measuring and displaying gene expression in samples derived
CC     from the human heart via microarrays. By measuring gene expression, the
CC     probes are useful for predicting, diagnosing, grading, staging,
CC     monitoring and prognosing diseases of the human heart and vascular system
CC     e.g. cardiovascular disease; hypertension, cardiac arrhythmias and
CC     congenital heart disease.
CC     Note: The sequence data for this patent did not form part of the printed
CC     specification, but was obtained in electronic format directly from WIPO
CC     at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ     Sequence 527 BP; 175 A; 99 C; 110 G; 143 T; 0 other;

Query Match     5.0%; Score 19; DB 22; Length 527;
Best Local Similarity 100.0%; Pred. No. 9;
Matches    19; Conservative    0; Mismatches    0; Indels    0; Gaps    0;
```

```
Qy    162 TGTTCTTTGGAATAAAAAC 180
          |||||||||||||||||||
Db    477 TGTTCTTTGGAATAAAAAC 495
```

RESULT 8
AAK09951
ID   AAK09951 standard; DNA; 527 BP.
XX
AC   AAK09951;
XX
DT   05-NOV-2001 (first entry)
XX
DE   Human brain expressed single exon probe SEQ ID NO: 9942.
XX
KW   Human; brain expressed exon; gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0235359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains -
XX
PS   Example 4; SEQ ID NO: 9942; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is one of the probes of the
CC   invention.
XX
SQ   Sequence 527 BP; 175 A; 99 C; 110 G; 143 T; 0 other;

Query Match   5.0%; Score 19; DB 22; Length 527;
Best Local Similarity 100.0%;
Matches   19; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    162 TGTTCTTTGGAATAAAAAC 180
          |||||||||||||||||||
Db    477 TGTTCTTTGGAATAAAAAC 495
```

RESULT 9
AAK35845
ID   AAK35845 standard; DNA; 527 BP.
XX
AC   AAK35845;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human bone marrow expressed single exon probe SEQ ID NO: 10402.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0235359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
XX
PS   Example 4; SEQ ID NO: 10402; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC   the probes of the invention.
XX
SQ   Sequence 527 BP; 175 A; 99 C; 110 G; 143 T; 0 other;

Query Match   5.0%; Score 19; DB 22; Length 527;
Best Local Similarity 100.0%;
Matches   19; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    162 TGTTCTTTGGAATAAAAAC 180
          |||||||||||||||||||
Db    477 TGTTCTTTGGAATAAAAAC 495
```

RESULT 10
AAI17155
ID   AAI17155 standard; DNA; 527 BP.
XX
AC   AAI17155;
XX
DT   12-OCT-2001 (first entry)
XX
DE   Probe #7088 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.

DR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 7088; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 527 BP; 175 A; 99 C; 110 G; 143 T; 0 other;

Qy   162 TGTTCTTTGGAATAAAAAC 180
         |||||||||||||||||||
Db   477 TGTTCTTTGGAATAAAAAC 495

Query Match   100.0%; Score 19; DB 22; Length 527;
Best Local Similarity   5.0%; Pred. No. 9;
Matches   19; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

DT   17-OCT-2001 (first entry)
XX
DE   Probe #10248 used to measure gene expression in human placenta sample.
XX
AC   AAI41562;
XX
ID   AAI41562 standard; DNA; 527 BP.
XX
RESULT 11
AAI41562

Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;

DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 10248; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 527 BP; 175 A; 99 C; 110 G; 143 T; 0 other;

Qy   162 TGTTCTTTGGAATAAAAAC 180
         |||||||||||||||||||
Db   477 TGTTCTTTGGAATAAAAAC 495

Query Match   100.0%; Score 19; DB 22; Length 527;
Best Local Similarity   5.0%; Pred. No. 9;
Matches   19; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

DT   19-AUG-2002 (first entry)
XX
DE   Human genome-derived single exon probe from lung SEQ ID No 10069.
XX
AC   ABS10078;
XX
ID   ABS10078 standard; DNA; 527 BP.
XX
RESULT 12
ABS10078

Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW   chronic obstructive pulmonary disease; interstitial lung disease;
KW   familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW   tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW   Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW   pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW   pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW   primary ciliary dyskinesis; pulmonary hypertension;
KW   hyaline membrane disease.
XX
OS   Homo sapiens.
XX
PN   WO200186003-A2.
XX
PD   15-NOV-2001.
XX
PF   30-JAN-2001; 2001WO-US00665.
XX
PR   04-FEB-2000; 2000US-180312P.
PR   26-MAY-2000; 2000US-207456P.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-234687P.
PR   27-SEP-2000; 2000US-236359P.
PR   04-OCT-2000; 2000US-2363 59P.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2002-111183/15.
XX
PT   Spatially-addressable set of single exon nucleic acid probes, used to
PT   measure gene expression in human lung samples -
XX
PS   Claim 1; SEQ ID No 10069; 634pp; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived

```
CC   from human lung comprising single exon nucleic acid probes having one of
CC   12614 nucleic acid sequences mentioned in the specification, or their
CC   complements or the 12387 open reading frames derived from the 12614
CC   probes. Also included are a microarray comprising the novel set of
CC   probes; the novel set of probes which hybridise at high stringency to a
CC   nucleic acid expressed in the human lung; measuring gene expression in a
CC   sample derived from human lung, comprising (a) contacting the array with
CC   a collection of detectably labeled nucleic acids derived from human lung
CC   mRNA, and (b) measuring the label detectably bound to each probe of
CC   the array; identifying exons in a eukaryotic genome, comprising
CC   (a) algorithmically predicting at least one exon from genomic sequences
CC   of the eukaryote; and (b) detecting specific hybridisation of detectably
CC   labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC   having a fragment identical to the predicted exon, the probe is included
CC   in the above mentioned microarray; assigning exons to a single gene,
CC   comprising (a) identifying exons from genomic sequence by the method
CC   above and (b) measuring the expression of each of the exons in several
CC   tissues and/or cell types using hybridisation to a single exon
CC   microarrays having a probe with the exon, where a common pattern of
CC   expression of the exons in the tissues and/or cell types indicates that
CC   the exons should be assigned to a single gene; a peptide comprising one
CC   of 12011 sequences, mentioned in the specification, or encoded by the
CC   probes/open reading frames (ORF). The probes are used for gene
CC   expression analysis, and for identifying exons in a gene, particularly
CC   using human lung derived mRNA and for the study of lung diseases
CC   such as asthma, lung cancer, chronic obstructive pulmonary disease
CC   (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC   fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC   Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC   haemosiderosis, pulmonary histiocytosis, lymphangioleiomyontosis,
CC   pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC   pulmonary dysplasia, primary ciliary dyskinesis, pulmonary hypertension
CC   and hyaline membrane disease. The present sequence is a single exon
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 527 BP; 175 A; 99 C; 110 G; 143 T; 0 other;

Qy    162 TGTTCTTTGGAATAAAAAC 180
          |||||||||||||||||||
Db    477 TGTTCTTTGGAATAAAAAC 495

Query Match             5.0%;  Score 19;  DB 24;  Length 527;
Best Local Similarity   100.0%;  Pred. No. 9;
Matches    19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

RESULT 13
ABA17906
ID   ABA17906 standard; DNA; 9929 BP.
XX
AC   ABA17906;
XX
DT   23-JAN-2002 (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 10237.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiarrhythmic;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180629.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   16-MAR-2000; 2000US-0190076.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0216980.
PR   11-JUL-2000; 2000US-0217487.
PR   14-JUL-2000; 2000US-0217496.
PR   26-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226679.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   23-AUG-2000; 2000US-0227182.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
```

PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
PS   Disclosure; SEQ ID NO 10237; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone,
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 9929 BP; 2748 A; 1678 C; 1979 G; 3524 T; 0 other;

Qy   128 TTAAGCCCCTGTTCTTTTTC 146
              ||||||||||||||||||||
Db   7124 TTAAGCCCCTGTTCTTTTTC 7142

Query Match   5.0%; Score 19; DB 22; Length 9929;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 14
AAK58913/c
ID   AAK58913 standard; cDNA; 249 BP.
XX
AC   AAK58913;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:3973.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO20015718-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.

PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236870.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0242221.
PR   20-OCT-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   08-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250391.
PR   01-DEC-2000; 2000US-0250391.
PR   01-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251160.
PR   05-DEC-2000; 2000US-0251479.
PR   06-DEC-2000; 2000US-0251479.
PR   05-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI: 2001-541565/60.

```
PR   26-JUL-2000;  2000US-0220963.
PR   26-JUL-2000;  2000US-0220964.
PR   14-AUG-2000;  2000US-0224518.
PR   14-AUG-2000;  2000US-0224519.
PR   14-AUG-2000;  2000US-0225213.
PR   14-AUG-2000;  2000US-0225214.
PR   14-AUG-2000;  2000US-0225266.
PR   14-AUG-2000;  2000US-0225267.
PR   14-AUG-2000;  2000US-0225268.
PR   14-AUG-2000;  2000US-0225270.
PR   14-AUG-2000;  2000US-0225447.
PR   14-AUG-2000;  2000US-0225757.
PR   14-AUG-2000;  2000US-0225758.
PR   14-AUG-2000;  2000US-0225759.
PR   18-AUG-2000;  2000US-0226279.
PR   22-AUG-2000;  2000US-0226681.
PR   22-AUG-2000;  2000US-0226868.
PR   22-AUG-2000;  2000US-0227182.
PR   23-AUG-2000;  2000US-0227009.
PR   30-AUG-2000;  2000US-0228924.
PR   01-SEP-2000;  2000US-0229287.
PR   01-SEP-2000;  2000US-0229343.
PR   01-SEP-2000;  2000US-0229344.
PR   01-SEP-2000;  2000US-0229345.
PR   05-SEP-2000;  2000US-0229509.
PR   05-SEP-2000;  2000US-0229513.
PR   06-SEP-2000;  2000US-0230437.
PR   06-SEP-2000;  2000US-0230438.
PR   01-SEP-2000;  2000US-0231242.
PR   01-SEP-2000;  2000US-0231243.
PR   08-SEP-2000;  2000US-0231244.
PR   08-SEP-2000;  2000US-0231413.
PR   08-SEP-2000;  2000US-0231414.
PR   08-SEP-2000;  2000US-0231423.
PR   08-SEP-2000;  2000US-0232080.
PR   08-SEP-2000;  2000US-0232081.
PR   12-SEP-2000;  2000US-0231968.
PR   14-SEP-2000;  2000US-0232397.
PR   14-SEP-2000;  2000US-0232398.
PR   14-SEP-2000;  2000US-0232399.
PR   14-SEP-2000;  2000US-0232400.
PR   14-SEP-2000;  2000US-0232401.
PR   14-SEP-2000;  2000US-0233063.
PR   14-SEP-2000;  2000US-0233064.
PR   14-SEP-2000;  2000US-0233065.
PR   21-SEP-2000;  2000US-0234223.
PR   21-SEP-2000;  2000US-0234274.
PR   25-SEP-2000;  2000US-0234997.
PR   25-SEP-2000;  2000US-0234998.
PR   26-SEP-2000;  2000US-0235484.
PR   27-SEP-2000;  2000US-0235834.
PR   27-SEP-2000;  2000US-0235836.
PR   29-SEP-2000;  2000US-0236327.
PR   29-SEP-2000;  2000US-0236367.
PR   29-SEP-2000;  2000US-0236368.
PR   29-SEP-2000;  2000US-0236369.
PR   29-SEP-2000;  2000US-0236370.
PR   02-OCT-2000;  2000US-0236802.
PR   02-OCT-2000;  2000US-0237037.
PR   02-OCT-2000;  2000US-0237038.
PR   02-OCT-2000;  2000US-0237039.
PR   02-OCT-2000;  2000US-0237040.
PR   13-OCT-2000;  2000US-0239935.
PR   13-OCT-2000;  2000US-0239937.
PR   20-OCT-2000;  2000US-0240960.
PR   20-OCT-2000;  2000US-0241221.
PR   20-OCT-2000;  2000US-0241785.
PR   20-OCT-2000;  2000US-0241786.
PR   20-OCT-2000;  2000US-0241787.
PR   20-OCT-2000;  2000US-0241808.
PR   20-OCT-2000;  2000US-0241809.
PR   20-OCT-2000;  2000US-0241826.
PR   01-NOV-2000;  2000US-0244617.
PR   08-NOV-2000;  2000US-0246474.

PR   08-NOV-2000;  2000US-0246475.
PR   08-NOV-2000;  2000US-0246476.
PR   08-NOV-2000;  2000US-0246477.
PR   08-NOV-2000;  2000US-0246478.
PR   08-NOV-2000;  2000US-0246523.
PR   08-NOV-2000;  2000US-0246524.
PR   08-NOV-2000;  2000US-0246525.
PR   08-NOV-2000;  2000US-0246526.
PR   08-NOV-2000;  2000US-0246527.
PR   08-NOV-2000;  2000US-0246528.
PR   08-NOV-2000;  2000US-0246532.
PR   08-NOV-2000;  2000US-0246609.
PR   08-NOV-2000;  2000US-0246610.
PR   08-NOV-2000;  2000US-0246611.
PR   08-NOV-2000;  2000US-0246613.
PR   17-NOV-2000;  2000US-0249207.
PR   17-NOV-2000;  2000US-0249208.
PR   17-NOV-2000;  2000US-0249209.
PR   17-NOV-2000;  2000US-0249210.
PR   17-NOV-2000;  2000US-0249211.
PR   17-NOV-2000;  2000US-0249212.
PR   17-NOV-2000;  2000US-0249213.
PR   17-NOV-2000;  2000US-0249214.
PR   17-NOV-2000;  2000US-0249215.
PR   17-NOV-2000;  2000US-0249216.
PR   17-NOV-2000;  2000US-0249217.
PR   17-NOV-2000;  2000US-0249218.
PR   17-NOV-2000;  2000US-0249244.
PR   17-NOV-2000;  2000US-0249245.
PR   17-NOV-2000;  2000US-0249264.
PR   17-NOV-2000;  2000US-0249265.
PR   17-NOV-2000;  2000US-0249297.
PR   17-NOV-2000;  2000US-0249299.
PR   17-NOV-2000;  2000US-0249300.
PR   01-DEC-2000;  2000US-0250160.
PR   01-DEC-2000;  2000US-0250591.
PR   05-DEC-2000;  2000US-0250991.
PR   05-DEC-2000;  2000US-0251030.
PR   05-DEC-2000;  2000US-0251479.
PR   05-DEC-2000;  2000US-0251856.
PR   05-DEC-2000;  2000US-0251988.
PR   08-DEC-2000;  2000US-0251868.
PR   08-DEC-2000;  2000US-0251869.
PR   08-DEC-2000;  2000US-0251989.
PR   08-DEC-2000;  2000US-0251990.
PR   08-DEC-2000;  2000US-0254097.
PR   11-DEC-2000;  2000US-0255719.
PR   05-JAN-2001;  2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI: 2001-483426/52.
DR   P-PSDB; AAM86132.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Claim 1; SEQ ID NO 3973; 307lpp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
```

```
CC  diagnose and treat immune/haematopoietic-related diseases, especially
CC  cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC  to AAK87694 represent human immune/haematopoietic antigen genomic
CC  sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC  represent sequences used in the exemplification of the present invention.
XX
SQ  Sequence 249 BP; 68 A; 64 C; 40 G; 74 T; 3 other;
XX
Qy  257 GGGATGGTGATGGAGAGATG 274
        ||||||||||||||||||||
Db   78 GGGATGGTGATGGAGAGATG 61

Query Match             4.8%;  Score 18;  DB 22;  Length 249;
Best Local Similarity   100.0%;  Pred. No. 30;
Matches   18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

RESULT 15
AAK74060/c
ID   AAK74060 standard; DNA; 333 BP.
XX
AC   AAK74060;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:28872.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   14-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   18-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
```

```
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229945.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232081.
PR   14-SEP-2000; 2000US-0232080.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   21-SEP-2000; 2000US-0233065.
PR   14-SEP-2000; 2000US-0234223.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236227.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   01-NOV-2000; 2000US-0241826.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0245527.
PR   08-NOV-2000; 2000US-0245528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
```

```
PR   17-NOV-2000;   2000US-0249211.
PR   17-NOV-2000;   2000US-0249212.
PR   17-NOV-2000;   2000US-0249213.
PR   17-NOV-2000;   2000US-0249214.
PR   17-NOV-2000;   2000US-0249215.
PR   17-NOV-2000;   2000US-0249916.
PR   17-NOV-2000;   2000US-0249217.
PR   17-NOV-2000;   2000US-0249218.
PR   17-NOV-2000;   2000US-0249244.
PR   17-NOV-2000;   2000US-0249245.
PR   17-NOV-2000;   2000US-0249264.
PR   17-NOV-2000;   2000US-0249265.
PR   17-NOV-2000;   2000US-0249297.
PR   17-NOV-2000;   2000US-0249299.
PR   17-NOV-2000;   2000US-0249900.
PR   01-DEC-2000;   2000US-0250160.
PR   01-DEC-2000;   2000US-0250391.
PR   05-DEC-2000;   2000US-0251030.
PR   05-DEC-2000;   2000US-0251988.
PR   05-DEC-2000;   2000US-0256719.
PR   06-DEC-2000;   2000US-0251479.
PR   08-DEC-2000;   2000US-0251856.
PR   08-DEC-2000;   2000US-0251868.
PR   08-DEC-2000;   2000US-0251869.
PR   08-DEC-2000;   2000US-0251989.
PR   08-DEC-2000;   2000US-0251990.
PR   11-DEC-2000;   2000US-0254097.
PR   05-JAN-2001;   2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis
XX
PS   Disclosure; SEQ ID NO 28872; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I).
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 333 BP; 84 A; 91 C; 64 G; 94 T; 0 other;

Query Match             4.8%;  Score 18;  DB 22;  Length 333;
Best Local Similarity   100.0%;
Matches    18;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          257 GGGATGGTGATGGAGATG 274
                ||||||||||||||||||
Db          145 GGGATGGTGATGGAGATG 128
```

us-09-803-719-222.oli.rge

OM nucleic - nucleic search, using sw model

Run on: January 31, 2003, 02:17:29 ; Search time 2723 Seconds
(without alignments)
4039.978 Million cell updates/sec

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

Title: US-09-803-719-222
Perfect score: 378
Sequence: 1 tacggctgcttaagacgact.........taccattggagaacatgcg 378
Scoring table: OLIGO_NUC
                Gapop 60.0 , Gapext 60.0
Searched: 2054640 seqs, 1455102878 residues
Word size :   0
Total number of hits satisfying chosen parameters: 4109280
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database :   GenEmbl:*

```
 1:  gb_ba:*
 2:  gb_htg:*
 3:  gb_in:*
 4:  gb_om:*
 5:  gb_ov:*
 6:  gb_pat:*
 7:  gb_ph:*
 8:  gb_pl:*
 9:  gb_pr:*
10:  gb_ro:*
11:  gb_ro:*
12:  gb_sts:*
13:  gb_sy:*
14:  gb_un:*
15:  gb_vi:*
16:  em_ba:*
17:  em_fun:*
18:  em_hum:*
19:  em_in:*
20:  em_mu:*
21:  em_om:*
22:  em_or:*
23:  em_ov:*
24:  em_pat:*
25:  em_ph:*
26:  em_pl:*
27:  em_ro:*
28:  em_sts:*
29:  em_un:*
30:  em_vi:*
31:  em_htg_hum:*
32:  em_htg_inv:*
33:  em_htg_mus:*
34:  em_htg_other:*
35:  em_htg_pln:*
36:  em_htg_rod:*
37:  em_htg_mam:*
38:  em_htg_vrt:*
39:  em_sy:*
40:  em_htgo_hum:*
41:  em_htgo_mus:*
     em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		378	100.0	378	6	AX245292	AX245292 Sequence
2		238	63.0	322	6	AX245365	AX245365 Sequence
3		238	63.0	327	6	AX245346	AX245346 Sequence
4		232	61.4	318	6	AX245370	AX245370 Sequence
5		119	31.5	59231	9	AL135937	AL135937 Human DNA
6	c	20	5.3	65118	2	AC084187	AC084187 Homo sapi
7	c	20	5.3	65811	9	AL451145	AL451145 Human DNA
8	c	20	5.3	100403	9	AC003087	AC003087 Homo sapi
9	c	20	5.3	122370	2	AP004044	AP004044 Homo sapi
10		20	5.3	199591	2	AP005251	AP005251 Oryza sat
11		20	5.3	131682	9	AL672277	AL672277 Human DNA
12	c	20	5.3	132257	9	AC004520	AC004520 Homo sapi
13	c	20	5.3	154469	2	AC046149	AC046149 Mus muscu
14		20	5.3	159108	10	AC009295	AC009295 Mus muscu
15	c	20	5.3	162394	2	AC119775	AC119775 Rattus no
16		20	5.3	171371	9	AP001132	AP001132 Homo sapi
17		20	5.3	171638	10	AC124708	AC124708 Mus muscu
18		20	5.3	187983	2	AC026210	AC026210 Homo sapi
19		20	5.3	187987	2	AC009639	AC009639 Homo sapi
20	c	20	5.3	191141	2	AC069374	AC069374 Homo sapi
21	c	20	5.3	202251	2	AC104472	AC104472 Homo sapi
22	c	20	5.3	268444	2	AC110152	AC110152 Homo sapi
23		19	5.0	6469	9	AB007869	AB007869 Homo sapi
24	c	19	5.0	6854	1	BSU39230	BSU39230 Bacillus sp
25	c	19	5.0	7260	9	AF339163	AF339163 Cynocepha
26	c	19	5.0	56703	5	AC100286	AC100286 Mus muscu
27		19	5.0	60648	5	AL672044	AL672044 Zebrafish
28		19	5.0	9939	1	AE006471	AE006471 Salmonell
29		19	5.0	106607	9	AL365259	AL365259 Human DNA
30	c	19	5.0	113810	2	AC068292	AC068292 Homo sapi
31		19	5.0	130468	2	AL831801	AL831801 Oryza sat
32	c	19	5.0	141818	2	AL133389	AL133389 Homo sapi
33		19	5.0	142522	9	AC090625	AC090625 Homo sapi
34		19	5.0	155132	2	AC095530	AC095530 Rattus no
35	c	19	5.0	158475	9	AC097463	AC097463 Homo sapi
36		19	5.0	161731	2	AC128360	AC128360 Rattus no
37	c	19	5.0	164074	9	AC103987	AC103987 Homo sapi
38		19	5.0	168114	9	AL354828	AL354828 Human DNA
39	c	19	5.0	169886	2	AC016578	AC016578 Homo sapi
40	c	19	5.0	170334	9	AC016579	AC016579 Homo sapi
41		19	5.0	175395	2	AC119119	AC119119 Rattus no
42	c	19	5.0	177147	9	AC007539	AC007539 Homo sapi
43		19	5.0	181300	2	HSRA512N4	HSRA512N4 Human DNA
44		19	5.0	181413	2	AP001352	AP001352 Homo sapi
45		19	5.0	188132	9	AC079304	AC079304 Homo sapi

ALIGNMENTS

```
RESULT 1
AX245292
LOCUS       AX245292                378 bp    DNA    linear    PAT 28-SEP-2001
DEFINITION  Sequence 222 from Patent WO0166753.
ACCESSION   AX245292
VERSION     AX245292.1  GI:15859966
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 378)
AUTHORS     Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
            Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
```

```
TITLE        Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
             Stache-Crain,B.
             Human genes and gene expression products
JOURNAL      Patent: WO 0166753-A 222 13-SEP-2001;
             Chiron Corporation (US); Hyseq Inc. (US)
FEATURES             Location/Qualifiers
    source   1..378
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
BASE COUNT    100 a       76 c      101 g      101 t
ORIGIN

Query Match        100.0%;  Score 378;  DB 6;  Length 378;
Best Local Similarity 100.0%;  Pred. No. 4.9e-225;
Matches  378;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 TACGGGCTGCTTAAGACGACTTAAGGGGGGAAATGACGCAGCGGGCTCTTAGAGACGAACATATGG   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 TACGGGCTGCTTAAGACGACTTAAGGGGGGAAATGACGCAGCGGGCTCTTAGAGACGAACATATGG   60

Qy    61 AAAACACCCAACCCGGAGTCTCTCACAAGCTTGAATGTGTGGTTCTGGAGCTGAAGGAGTG  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AAAACACCCAACCCGGAGTCTCTCACAAGCTTGAATGTGTGGTTCTGGAGCTGAAGGAGTG  120

Qy   121 ACGGTTGTTAAGCCCCTGTTCTTTTCTTTGTTTAATCTAATGTTCTTTGGAATAAAAAAAC  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ACGGTTGTTAAGCCCCTGTTCTTTTCTTTGTTTAATCTAATGTTCTTTGGAATAAAAAAC  180

Qy   181 CTCCCTGCCCAAGTAGTACTTGGTTTTATGCTCAACATGCTTGACTGTTGAAAAGAGACC  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTCCCTGCCCAAGTAGTACTTGGTTTTATGCTCAACATGCTTGACTGTTGAAAAGAGACC  240

Qy   241 TTTGGCACACATTGAAGGGATGGTGATGGAGGATGCCAATCAGGTGGTGGCGCAG  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TTTGGCACACATTGAAGGGATGGTGATGGAGGATGCCAATCAGGTGGTGGCGCAG  300

Qy   301 CTATGTTGGTAGCTATAGCAGAGAGAGTCTTCTTCTTTGGCAAAGATTCCTCCCGGGAAGGAAGGTA  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CTATGTTGGTAGCTATTACCAGAGAAGTCTTCTTGGCAAAGATTCCTCCCGGGAAGGAAGGTA  360

Qy   361 CCAATTGGAGAACCATGCG  378
         |||||||||||||||||||
Db   361 CCAATTGGAGAACCATGCG  378
```

```
RESULT 2
LOCUS        AX245365                  322 bp        DNA      linear    PAT  28-SEP-2001
DEFINITION   Sequence 295 from Patent WO0166753.
ACCESSION    AX245365
VERSION      AX245365.1  GI:1586009
KEYWORDS
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 322)
  AUTHORS    Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
             Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
             Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
             Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
             Stache-Crain,B.
  TITLE      Human genes and gene expression products
  JOURNAL    Patent: WO 0166753-A 295 13-SEP-2001;
             Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES             Location/Qualifiers
    source   1..322
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
BASE COUNT     90 a       64 c       78 g       89 t        1 others
ORIGIN
```

```
Query Match        63.0%;  Score 238;  DB 6;  Length 322;
```

```
Best Local Similarity 100.0%;  Pred. No. 3.2e-137;
Matches  238;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    59 GGAAAAACACCCAAGCCGGAGTCTCTCACAAGCTTGAATGTGTGGTTCTGGAGCTGAAGGAT  118
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    67 GGAAAAACACCCAAGCCGGAGTCTCTCACAAGCTTGAATGTGTGGTTCTGGAGCTGAAGGAT  126

Qy   119 ACCTCCCTGCCAAGTAGTACTTGGTTTTATGCTCAACATGCTTGACTGTTGAAAAGAGA  178
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   127 GCACGGTTGTTAAGCCCCTGTTCTTTTCTTTGTTTAATCTAATGTTCTTTGGAATAAA  186

Qy   179 ACCTCCCTGCCAAGTAGTACTTGGTTTTATGCTCAACATGCTTGACTGTTGAAAAGAGA  238
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   187 ACCTCCCTGCCCAAGTAGTACTTGGTTTTATGCTCAACATGCTTGACTGTTGAAAAGAGA  246

Qy   239 CCTTTGGCACACATTGAAAGGGATGGTGATGGAGGATGCCAATCCATGGAATCAGGTGGC  296
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   247 CCTTTGGCACACATTGAAAGGGATGGTGATGGAGGATGCCAATCCATGGAATCAGGTGGC  304
```

```
RESULT 3
LOCUS        AX245346                  327 bp        DNA      linear    PAT  28-SEP-2001
DEFINITION   Sequence 276 from Patent WO0166753.
ACCESSION    AX245346
VERSION      AX245346.1  GI:15860020
KEYWORDS
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 327)
  AUTHORS    Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
             Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
             Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
             Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
             Stache-Crain,B.
  TITLE      Human genes and gene expression products
  JOURNAL    Patent: WO 0166753-A 276 13-SEP-2001;
             Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES             Location/Qualifiers
    source   1..327
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
BASE COUNT     92 a       64 c       81 g       90 t
ORIGIN
```

```
Query Match        63.0%;  Score 238;  DB 6;  Length 327;
Best Local Similarity 100.0%;  Pred. No. 3.2e-137;
Matches  238;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    59 GGAAAAACACCCAAGCCGGAGTCTCTCACAAGCTTGAATGTGTGGTTCTGGAGCTGAAGGAT  118
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    52 GGAAAAACACCCAAGCCGGAGTCTCTCACAAGCTTGAATGTGTGGTTCTGGAGCTGAAGGAT  111

Qy   119 GCACGGTTGTTAAGCCCCTGTTCTTTTCTTTGTTTAATCTAATGTTCTTTGGAATAAA  178
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   112 GCACGGTTGTTAAGCCCCTGTTCTTTTCTTTGTTTAATCTAATGTTCTTTGGAATAAA  171

Qy   179 ACCTCCCTGCCAAGTAGTACTTGGTTTTATGCTCAACATGCTTGACTGTTGAAAAGAGA  238
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   172 ACCTCCCTGCCAAGTAGTACTTGGTTTTATGCTCAACATGCTTGACTGTTGAAAAGAGA  231

Qy   239 CCTTTGGCACACATTGAAAGGGATGGTGATGGAGGATGCCAATCCATGGAATCAGGTGGC  296
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   232 CCTTTGGCACACATTGAAAGGGATGGTGATGGAGGATGCCAATCCATGGAATCAGGTGGC  289
```

```
RESULT 4
LOCUS        AX245370                  318 bp        DNA      linear    PAT  28-SEP-2001
DEFINITION   Sequence 300 from Patent WO0166753.
ACCESSION    AX245370
```

VERSION   AX245370.1 GI:15860044
KEYWORDS   human.
SOURCE   Homo sapiens
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 318)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
             Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
             Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
             Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
             Stache-Crain,B.
  TITLE   Human genes and gene expression products
  JOURNAL   Patent: WO 0166753-A 300 13-SEP-2001;
             Chiron Corporation (US); Hyseq Inc. (US)
FEATURES   Location/Qualifiers
     source   1..318
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT   90 a   63 c   76 g   89 t
ORIGIN

Qy   59 GGAAAACACCCAAGCCGGAGTCTCTCACAAGCTTCAAAGCTTGAATGTGTGTTCTCGGAGCTGAAGGAT 118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   67 GGAAAAACACCCAAGCCGGAGTCTCTCACAAGCTTCAAAGCTTGAATGTGTGTTCTCGGAGCTGAAGGAT 126

Qy   119 GCACGGTTGTTAAGCCCCCTGTCTCTTTTCCGTTGTTTAATCTAATGTGTGTTCTTGGAATAAAA 178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   127 GCACGGTTGTTAAGCCCCCTGTCTCTTTTCCGTTGTTTAATCTAATGTGTGTTCTTGGAATAAAA 186

Qy   179 ACCTCCCTGCCCAAGTAGTACTTGGTTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGA 238
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   187 ACCTCCCTGCCCAAGTAGTACTTGGTTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGA 246

Qy   239 CCTTTGGGCACACATTGAAGGAGTGGGTGATGGAGATGCCAAATCCATGGAATCA 290
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   247 CCTTTGGGCACACATTGAAGGAGTGGGTGATGGAGATGCCAAATCCATGGAATCA 298

RESULT 5
LOCUS   AL135937   59231 bp   DNA   linear   PRI 15-MAR-2001
DEFINITION Human DNA sequence from clone RP1-278022 on chromosome 20. Contains
            STSs and GSSs, complete sequence.
ACCESSION   AL135937
VERSION   AL135937.22 GI:7248202
KEYWORDS   HTG.
SOURCE   Homo sapiens
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 59231)
  AUTHORS   Phillimore,B.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT   On Mar 15, 2000 this sequence version replaced gi:7242336.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence

Best Local Similarity   61.4%;
Matches 232; Conservative 0; Pred. No. 1.8e-133;
Query Match 100.0%; Score 232; DB 6; Length 318;
Mismatches 0; Indels 0; Gaps 0;

FEATURES   Location/Qualifiers
     source   1..59231
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP1-278022"
                     /clone_lib="RPCI-1"
     repeat_region   58..213
                     /note="MER5A repeat: matches 9..188 of consensus"
     repeat_region   233..409
                     /note="MER3 repeat: matches 6..191 of consensus"
     repeat_region   572..795
                     /note="MIR repeat: matches 3..241 of consensus"
     repeat_region   2927..2992
                     /note="Alu repeat: matches 242..307 of consensus"
     repeat_region   3019..3635
                     /note="L1ME2 repeat: matches 5541..6154 of consensus"
     repeat_region   3880..4411
                     /note="L1M4 repeat: matches 4015..4580 of consensus"
     repeat_region   4412..4600
                     /note="MER58A repeat: matches 32..224 of consensus"
     misc_feature   complement(5199..5709)
                     /note="match: GSS: Em:AQ613263"
     repeat_region   7115..7585
                     /note="Charlie4a repeat: matches 25..507 of consensus"
     repeat_region   7631..7695
                     /note="MIR repeat: matches 13..81 of consensus"
     repeat_region   8546..9223
                     /note="L1MC4 repeat: matches 7117..7739 of consensus"
     repeat_region   9986..10209
                     /note="AluSq repeat: matches 1..225 of consensus"
     repeat_region   10225..10250
                     /note="13 copies 2 mer aa 96% conserved"
     misc_feature   complement(10684..11081)
                     /note="match: GSS: Em:AQ662878"
     repeat_region   11093..11168
                     /note="38 copies 2 mer aa 72% conserved"
     repeat_region   11575..12292
                     /note="L1MA9 repeat: matches 5545..6278 of consensus"
     misc_feature   12860..13177
                     /note="match: GSS: Em:AQ035618"
     repeat_region   13327..13423
                     /note="L2 repeat: matches 2657..2750 of consensus"
     repeat_region   13719..13813
                     /note="L2 repeat: matches 2195..2293 of consensus"
     repeat_region   13814..14087
                     /note="L1MA5 repeat: matches 6025..6298 of consensus"
     repeat_region   14088..14219
                     /note="L2 repeat: matches 2078..2195 of consensus"
     repeat_region   14965..15067
                     /note="MIR repeat: matches 169..262 of consensus"
     misc_feature   complement(15295..15803)
                     /note="match: GSS: Em:AQ754141"

was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
IMPORTANT: This sequence is not the entire insert of clone
RP1-278022 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP1-278022 is at 59132 in this sequence.
The true right end of clone RP5-1099D15 is at 100 in this sequence.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. RP1-278022 is from
the library RPCI-1 constructed by the group of Pieter de Jong. For
further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2.

```
misc_feature    15849..16248
                /note="match: GSS: Em:AQ694331"
repeat_region   15961..16498
                /note="MIR repeat: matches 5427..5964 of consensus"
repeat_region   16499..16808
                /note="L1ME3A repeat: matches 1..310 of consensus"
repeat_region   16809..17004
                /note="AluSx repeat: matches 5964..6159 of consensus"
repeat_region   17356..17544
                /note="L1ME3A repeat: matches 1..189 of consensus"
repeat_region   18267..18357
                /note="MIR repeat: matches 26..230 of consensus"
repeat_region   18358..18690
                /note="MIR repeat: matches 50..139 of consensus"
repeat_region   18691..18717
                /note="MER5B repeat: matches 2..341 of consensus"
repeat_region   18885..18952
                /note="MIR repeat: matches 24..50 of consensus"
misc_feature    complement(19364..19600)
                /note="34 copies 2 mer ga 79% conserved"
repeat_region   22132..22242
                /note="LTR33 repeat: matches 410..521 of consensus"
repeat_region   22243..22551
                /note="AluSx repeat: matches 1..310 of consensus"
repeat_region   22552..22951
                /note="LTR33 repeat: matches 17..410 of consensus"
repeat_region   23356..23643
                /note="AluSg repeat: matches 1..288 of consensus"
repeat_region   24124..24510
                /note="LTR16C repeat: matches 1..387 of consensus"
repeat_region   24747..24810
                /note="L1MC/D repeat: matches 5290..5579 of consensus"
misc_feature    24967..25054
                /note="32 copies 2 mer ta 68% conserved"
repeat_region   25134..25643
                /note="L1M1 repeat: matches 5678..5763 of consensus"
repeat_region   25767..26069
                /note="AluSq repeat: matches 1..313 of consensus"
repeat_region   26295..26579
                /note="AluSq repeat: matches 1..304 of consensus"
misc_feature    27383..27481
                /note="56 copies 2 mer ta 73% conserved"
repeat_region   28338..28469
                /note="MIR repeat: matches 20..142 of consensus"
repeat_region   28472..28583
                /note="L1MC4 repeat: matches 7841..7973 of consensus"
repeat_region   28584..28970
                /note="L1MC4 repeat: matches 7430..7849 of consensus"
repeat_region   30290..30876
                /note="L1MC4 repeat: matches 2124..2658 of consensus"
repeat_region   30877..31223
                /note="L2 repeat: matches 2124..2658 of consensus"
repeat_region   31224..31421
                /note="MER1B repeat: matches 1..337 of consensus"
repeat_region   31422..31739
                /note="L2 repeat: matches 1945..2124 of consensus"
repeat_region   31750..32046
                /note="AluY repeat: matches 1..309 of consensus"
repeat_region   32047..32188
                /note="AluSg repeat: matches 3..298 of consensus"
repeat_region   32189..32474
                /note="L2 repeat: matches 1809..1937 of consensus"
repeat_region   32475..32713
                /note="AluJo repeat: matches 1..293 of consensus"
repeat_region   34869..34938
                /note="L2 repeat: matches 1420..1809 of consensus"
repeat_region   35160..35351
                /note="MIR repeat: matches 191..261 of consensus"
repeat_region   36121..36273
                /note="MER5A repeat: matches 7..189 of consensus"
repeat_region   37904..37941
                /note="MIR repeat: matches 66..225 of consensus"
repeat_region   38788..38840
                /note="19 copies 2 mer aa 81% conserved"
```

```
repeat_region   38841..39254
                /note="MIR repeat: matches 37..84 of consensus"
repeat_region   39255..39304
                /note="MSTB repeat: matches 1..426 of consensus"
repeat_region   39310..39339
                /note="MIR repeat: matches 84..150 of consensus"
repeat_region   39314..39717
                /note="L1M3d repeat: matches 251..280 of consensus"
repeat_region   39314..39990
                /note="L1M3e repeat: matches -879..779 of consensus"
repeat_region   39974..39990
                /note="L1MA3 repeat: matches 3735..4011 of consensus"
repeat_region   39992..40126
                /note="AluJo repeat: matches 2..134 of consensus"
repeat_region   40153..40379
                /note="L1M9 repeat: matches 3999..4226 of consensus"
repeat_region   40380..40675
                /note="AluJo repeat: matches 1..306 of consensus"
repeat_region   40676..42726
                /note="L1MA9 repeat: matches 4226..6308 of consensus"
repeat_region   43333..43524
                /note="LTR16C repeat: matches 165..386 of consensus"
misc_feature    complement(43404..44004)
                /note="match: GSS: Em:AQ800109"
repeat_region   44765..44818
                /note="MIR repeat: matches 200..250 of consensus"
misc_feature    complement(45503..46337)
```

RESULT 6
AC084187

```
LOCUS       AC084187    65118 bp    DNA    linear    HTG 14-OCT-2000
DEFINITION  Homo sapiens clone RP11-609K10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC084187
VERSION     AC084187.1  GI:10801415
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 65118)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Unpublished
JOURNAL     
REFERENCE   2  (bases 1 to 65118)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campoplano,A., Castle,A.,
            Chopel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Illev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mhova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
```

```
Qy     148 GTTGTTTAATCTAATTGTTCTTTGGAATAAAAAAACCTCCCTGCCCAAGTAGTACTTGGTTTTA  207
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18738 GTTGTTTAATCTAATTGTTCTTTGGAATAAAAAAACCTCCCTGCCCAAGTAGTACTTGGTTTTA  18797

Qy     208 TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACAATTGAAGGGATGGGTGA  266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18798 TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACAATTGAAGGGATGGGTGA  18856
```

Query Match 31.5%; Score 119; DB 9; Length 59231;
Best Local Similarity 100.0%; Pred. No. 1.6e-62;
Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
          Tirrell,A., Travers,M., Trigillo,J., Vassiliev,H., Viel,R., Vo,A.,
          Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
          Zimmer,A. and Zody,M.
TITLE     Direct Submission
JOURNAL   Submitted (14-OCT-2000) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   All repeats were identified using RepeatMasker:
          Smit,A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          -------------------------------
          Center: Whitehead Institute/ MIT Center for Genome Research
          Center code: WIBR
          Web site: http://www-seq.wi.mit.edu
          Contact: sequence_submissions@genome.wi.mit.edu
          ------------------ Genome Center
          Center project name: L11460
          Center clone name: 609_K_10
          ------------------ Project Information

        * NOTE: This record contains 77 individual
        * sequencing reads that have not been assembled into
        * contigs. Runs of N are used to separate the reads
        * and the order in which they appear is completely
        * arbitrary. Low-pass sequence sampling is useful for
        * identifying clones that may be gene-rich and allows
        * overlap relationships among clones to be deduced.
        * However, it should not be assumed that this clone
        * will be sequenced to completion. In the event that
        * the record is updated, the accession number will
        * be preserved.

        *         1
        *       747: contig of 746 bp in length
        *       846: gap of 100 bp
        *       847
        *      1605: gap of 100 bp
        *      1606: contig of 759 bp in length
        *      1705: gap of 100 bp
        *      1706
        *      2465: contig of 760 bp in length
        *      2565: gap of 100 bp
        *      2566
        *      3313: contig of 748 bp in length
        *      3413: gap of 100 bp
        *      3414
        *      4145: contig of 732 bp in length
        *      4245: gap of 100 bp
        *      4246
        *      4953: contig of 708 bp in length
        *      5053: gap of 100 bp
        *      5054
        *      5818: contig of 765 bp in length
        *      5918: gap of 100 bp
        *      5919
        *      6664: contig of 746 bp in length
        *      6764: gap of 100 bp
        *      6765
        *      7496: contig of 732 bp in length
        *      7596: gap of 100 bp
        *      7597
        *      8344: contig of 748 bp in length
        *      8444: gap of 100 bp
        *      8445
        *      9198: contig of 754 bp in length
        *      9298: gap of 100 bp
        *      9299
        *     10047: contig of 749 bp in length
        *     10147: gap of 100 bp
        *     10148
        *     10902: contig of 755 bp in length
        *     11002: gap of 100 bp
        *     11003
        *     11785: contig of 783 bp in length
        *     11885: gap of 100 bp
        *     11886
        *     12636: contig of 751 bp in length
        *     12736: gap of 100 bp
        *     12737
        *     13503: contig of 767 bp in length
        *     13603: gap of 100 bp
        *     13604
        *     14357: contig of 754 bp in length
        *     14457: gap of 100 bp
        *     14458
        *     15208: contig of 751 bp in length
        *     15308: gap of 100 bp
        *     15309
        *     16026: contig of 718 bp in length
        *     16126: gap of 100 bp
        *     16127
        *     16867: contig of 741 bp in length
        *     16967: gap of 100 bp
        *     16968
        *     17697: contig of 730 bp in length
        *     17797: gap of 100 bp
        *     17798
        *     18538: contig of 741 bp in length
        *     18638: gap of 100 bp
        *     18639

        *     18639
        *     19403: contig of 765 bp in length
        *     19503: gap of 100 bp
        *     19504
        *     20248: contig of 745 bp in length
        *     20348: gap of 100 bp
        *     20349
        *     21101: contig of 753 bp in length
        *     21201: gap of 100 bp
        *     21202
        *     21932: contig of 731 bp in length
        *     22032: gap of 100 bp
        *     22033
        *     22790: contig of 758 bp in length
        *     22890: gap of 100 bp
        *     22891
        *     23641: contig of 751 bp in length
        *     23741: gap of 100 bp
        *     23742
        *     24480: contig of 739 bp in length
        *     24580: gap of 100 bp
        *     24581
        *     25321: contig of 741 bp in length
        *     25421: gap of 100 bp
        *     25422
        *     26169: contig of 748 bp in length
        *     26269: gap of 100 bp
        *     26270
        *     27016: contig of 747 bp in length
        *     27116: gap of 100 bp
        *     27117
        *     27881: contig of 765 bp in length
        *     27981: gap of 100 bp
        *     27982
        *     28747: contig of 766 bp in length
        *     28847: gap of 100 bp
        *     28848
        *     29607: contig of 760 bp in length
        *     29707: gap of 100 bp
        *     29708
        *     30457: contig of 750 bp in length
        *     30557: gap of 100 bp
        *     30558
        *     31317: contig of 760 bp in length
        *     31417: gap of 100 bp
        *     31418
        *     32167: contig of 750 bp in length
        *     32267: gap of 100 bp
        *     32268
        *     33014: contig of 747 bp in length
        *     33114: gap of 100 bp
        *     33115
        *     33850: contig of 736 bp in length
        *     33950: gap of 100 bp
        *     33951
        *     34698: contig of 748 bp in length
        *     34798: gap of 100 bp
        *     34799
        *     35541: contig of 743 bp in length
        *     35641: gap of 100 bp
        *     35642
        *     36392: contig of 751 bp in length
        *     36492: gap of 100 bp
        *     36493
        *     37242: contig of 750 bp in length
        *     37342: gap of 100 bp
        *     37343
        *     38091: contig of 749 bp in length
        *     38191: gap of 100 bp
        *     38192
        *     38923: contig of 732 bp in length
        *     39023: gap of 100 bp
        *     39024
        *     39777: contig of 754 bp in length
        *     39877: gap of 100 bp
        *     39878
        *     40651: contig of 774 bp in length
        *     40751: gap of 100 bp
        *     40752
        *     41468: contig of 717 bp in length
        *     41568: gap of 100 bp
        *     41569
        *     42287: contig of 719 bp in length
        *     42387: gap of 100 bp
        *     42388
        *     43133: contig of 746 bp in length
        *     43233: gap of 100 bp
        *     43234
        *     43970: contig of 737 bp in length
        *     44070: gap of 100 bp
        *     44071
        *     44814: contig of 744 bp in length
        *     44914: gap of 100 bp
        *     44915
        *     45672: contig of 758 bp in length
        *     45772: gap of 100 bp
        *     45773
        *     46509: contig of 737 bp in length
        *     46609: gap of 100 bp
        *     46610
        *     47344: contig of 735 bp in length
        *     47444: gap of 100 bp
        *     47445
        *     48184: contig of 740 bp in length
        *     48284: gap of 100 bp
        *     48285
        *     49047: contig of 763 bp in length
        *     49147: gap of 100 bp
        *     49148
        *     49884: contig of 737 bp in length
```

RESULT 7

AL451145/c

Query Match 5.3%; Score 20; DB 2; Length 65118;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     172 AATAAAAACCTCCCTGCCAA 191
           ||||||||||||||||||||
Db   53064 AATAAAAACCTCCCTGCCAA 53083
```

```
*  49885  49984: gap of 100 bp in length
*  49985  50737: contig of 753 bp in length
*  50837:        gap of 100 bp in length
*  50838  51165: contig of 728 bp in length
*  51566:        gap of 100 bp in length
*  51666  52397: contig of 732 bp in length
*  52498  52497: gap of 100 bp in length
*  52498  53249: contig of 752 bp in length
*  53350  53349: gap of 100 bp in length
*  53350  54084: contig of 735 bp in length
*  54085  54184: gap of 100 bp in length
*  54185  54930: contig of 746 bp in length
*  54931  55030: gap of 100 bp in length
*  55031  55767: contig of 737 bp in length
*  55768  55867: gap of 100 bp in length
*  55868  56655: contig of 788 bp in length
*  56656  56755: gap of 100 bp in length
*  56756  57500: contig of 745 bp in length
*  57501  57600: gap of 100 bp in length
*  57601  58357: contig of 757 bp in length
*  58358  58457: gap of 100 bp in length
*  58458  59200: contig of 743 bp in length
*  59201  59300: gap of 100 bp in length
```

```
LOCUS       AL451145    65811 bp    DNA     linear   PRI 08-JUN-2001
DEFINITION  Human DNA sequence from clone RP11-164A17 on chromosome 6, complete
            sequence.
ACCESSION   AL451145
VERSION     AL451145.15  GI:14348489
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 65811)
  AUTHORS   Garner,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-2001) Sanger Centre, Hinton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
  COMMENT   On Jun 11, 2001 this sequence version replaced gi:14268205.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-164A17 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-164A17 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP3-365O19 is at 65712 in this sequence.
            The true right end of clone RP11-545I5 is at 100 in this sequence.
FEATURES             Location/Qualifiers
     source          1..65811
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-164A17"
                     /clone_lib="RPCI-11.1"
     misc_feature    complement(1..114)
                     /note="match: GSS: Em:AQ744388"
     misc_feature    complement(1..100)
                     /note="match: GSS: Em:AQ351948"
     repeat_region   220..352
                     /note="match: MER5A repeat: matches 55..183 of consensus"
     repeat_region   2175..2344
                     /note="MER2 repeat: matches 2..165 of consensus"
     repeat_region   3016..3163
                     /note="L1PA7 repeat: matches 4691..6123 of consensus"
     repeat_region   3464..3765
                     /note="L2 repeat: matches 295..295 of consensus"
     misc_feature    6146..6787
                     /note="match: GSS: Em:AQ473725"
     repeat_region   3805..4497
                     /note="L2 repeat: matches 1242..1583 of consensus"
     repeat_region   4519..4552
                     /note="L2 repeat: matches 1713..2479 of consensus"
     misc_feature    complement(6757..7094)
                     /note="match: GSS: Em:B33281"
     repeat_region   7878..8157
                     /note="17 copies 2 mer aa 85% conserved"
     repeat_region   4555..5939
                     /note="AluSg repeat: matches 1..285 of consensus"
     repeat_region   10517..10648
                     /note="MER46A repeat: matches 1..134 of consensus"
     repeat_region   5994..6122
                     /note="L2 repeat: matches 2620..2750 of consensus"
     repeat_region   10814..11006
                     /note="MIT1H repeat: matches 327..525 of consensus"
     repeat_region   11139..11355
                     /note="MER58A repeat: matches 2..223 of consensus"
     repeat_region   11397..11750
                     /note="L1M4 repeat: matches 2580..2953 of consensus"
     misc_feature    11801..11976
                     /note="88 copies 2 mer ta 61% conserved"
     repeat_region   12076..12353
                     /note="L1MEc repeat: matches 1778..2060 of consensus"
     misc_feature    complement(12258..12657)
                     /note="match: GSS: Em:AQ076115"
     repeat_region   12429..12969
                     /note="L1MEc repeat: matches 1201..1743 of consensus"
     repeat_region   12981..13114
                     /note="MER46A repeat: matches 1..142 of consensus"
     repeat_region   13512..13792
                     /note="AluSg repeat: matches 1..304 of consensus"
     repeat_region   13308..14156
                     /note="MIR repeat: matches 5..262 of consensus"
     misc_feature    14242..14678
                     /note="match: GSS: Em:AQ825235"
     misc_feature    complement(15008..15441)
                     /note="match: GSS: Em:AQ391327"
                     complement(19964..20391)
                     /note="match: GSS: Em:AQ244751"
     repeat_region   20006..20049
                     /note="22 copies 2 mer at 77% conserved"
```

misc_feature   complement(20115..20578)
/note="match: GSS: Em:AQ067991"
misc_feature   complement(21112..21482)
/note="match: GSS: Em:AQ167816"
repeat_region   21335..21403
/note="match: GSS: Em:AQ167816"
repeat_region   22001..22050
/note="L2 repeat: matches 2608..2677 of consensus"
repeat_region   22077..22104
/note="MLT1B repeat: matches 1..50 of consensus"
repeat_region   22125..22222
/note="14 copies 2 mer ac 100% conserved"
repeat_region   23742..23817
/note="L1MA7 repeat: matches 6172..6283 of consensus"
repeat_region   24686..24831
/note="MER5A repeat: matches 114..188 of consensus"
repeat_region   25348..25438
/note="MER58 repeat: matches 9..174 of consensus"
repeat_region   27026..28100
/note="L1MB2 repeat: matches 6069..6161 of consensus"
repeat_region   28101..28370
/note="L1MEc repeat: matches 2378..3109 of consensus"
repeat_region   28371..28638
/note="AluSg repeat: matches 1..285 of consensus"
repeat_region   28623..28868
/note="L1MEc repeat: matches 2108..2378 of consensus"
repeat_region   29241..29270
/note="L1MEc repeat: matches 1259..1512 of consensus"
repeat_region   30792..30924
/note="15 copies 2 mer tg 86% conserved"
repeat_region   30925..31231
/note="AluJo repeat: matches 1..140 of consensus"
repeat_region   31232..31427
/note="AluSq repeat: matches 1..313 of consensus"
repeat_region   33067..33359
/note="AluJo repeat: matches 140..312 of consensus"
repeat_region   33714..34013
/note="AluJo repeat: matches 3..300 of consensus"
repeat_region   34130..34477
/note="AluSx repeat: matches 5..305 of consensus"
repeat_region   34804..35175
/note="L2 repeat: matches 1992..2327 of consensus"
repeat_region   35192..35325
/note="MSTD repeat: matches 1..392 of consensus"
repeat_region   35652..35807
/note="67 copies 2 mer aa 58% conserved"
repeat_region   39379..39450
/note="MER20 repeat: matches 43..217 of consensus"
repeat_region   39409..39460
/note="MIR repeat: matches 186..258 of consensus"
repeat_region   41006..41075
/note="L2 repeat: matches 2699..2750 of consensus"
repeat_region   41583..41807
/note="L2 repeat: matches 2668..2737 of consensus"
misc_feature   complement(42379..42834)
/note="AluJo repeat: matches 73..302 of consensus"
misc_feature   42825..43364
/note="match: GSS: Em:AQ182459"
repeat_region   45770..46464
/note="match: GSS: Em:AQ761068"
repeat_region   47303..47470
/note="L1PB3 repeat: matches 5461..6146 of consensus"
repeat_region   50450..50597
/note="L2 repeat: matches 2355..2522 of consensus"
repeat_region   50748..51019
/note="MIR repeat: matches 58..207 of consensus"
repeat_region   52473..53123
/note="L2 repeat: matches 2254..2534 of consensus"
repeat_region   53172..53433
/note="L1MEc repeat: matches 1026..1669 of consensus"
repeat_region   53511..55046
/note="L1MEc repeat: matches 1778..2103 of consensus"
repeat_region   55395..55441
/note="L1ME1 repeat: matches 4512..6160 of consensus"

repeat_region   55442..55564
/note="MER3 repeat: matches 156..201 of consensus"
repeat_region   55565..55869
/note="MERC repeat: matches 1..118 of consensus"
repeat_region   55870..56324
/note="AluY repeat: matches 1..303 of consensus"
repeat_region   56325..56474
/note="MERC repeat: matches 118..601 of consensus"
repeat_region   57024..57143
/note="MER3 repeat: matches 2..156 of consensus"
repeat_region   57330..57429
/note="ML1H repeat: matches 32..150 of consensus"
repeat_region   57535..57884
/note="MIR repeat: matches 53..155 of consensus"
repeat_region   57945..58113
/note="MER1B repeat: matches 1..337 of consensus"

Qy   45 TTAGAGGAAACAPATATGGAAAA 64
                |||||||||||||||||||||||
Db   23308 TTAGAGGAAACAPATATGGAAAA 23289

Query Match   5.3%; Score 20; DB 9; Length 65811;
Best Local Similarity 100.0%;
Matches   20; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 8
AP003087
LOCUS   AP003087   100403 bp   DNA   linear   PRI 20-DEC-2001
DEFINITION Homo sapiens genomic DNA, chromosome 11q clone:RP11-594014,
            complete sequences.
ACCESSION   AP003087
VERSION   AP003087.2 GI:17939957
KEYWORDS   HTG.
SOURCE   Homo sapiens DNA, clone:RP11-594014.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   Homo sapiens genomic DNA
JOURNAL   Published Only in Database (2001)
REFERENCE   2 (bases 1 to 100403)
AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (12-JAN-2001) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT   On Dec 19, 2001 this sequence version replaced gi:12231284.
FEATURES   Location/Qualifiers
     source   1..100403
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-594014"
BASE COUNT   30119 a 20315 c 19157 g 30812 t
ORIGIN

Query Match   5.3%; Score 20; DB 9; Length 100403;
Best Local Similarity 100.0%; Pred. No. 2;
Matches   20; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   155 AATCTAATGTTCTTTGGAAT 174
                ||||||||||||||||||||
Db   1892 AATCTAATGTTCTTTGGAAT 1911

RESULT 9

AP004044/c

```
LOCUS       AP004044       122370 bp    DNA    linear   HTG 21-MAR-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 8 clone
            clone:OJ1125_C01, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION   AP004044
VERSION     AP004044.1  GI:15208412
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group) DNA,
            clone:OJ1125_C01.
ORGANISM    Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 122370)
AUTHORS     Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE       Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, BAC
            clone:OJ1125_C01
JOURNAL     Published Only in Database (2001)
REFERENCE   2  (bases 1 to 122370)
AUTHORS     Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE       Direct Submission
JOURNAL     Submitted (15-AUG-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..122370
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="8"
                     /clone="OJ1125_C01"
BASE COUNT     36453 a  25424 c  26087 g  34354 t  52 others
ORIGIN

Qy          1 TACGGCTGCTTAAGACGACT 20
              ||||||||||||||||||||
Db     121735 TACGGCTGCTTAAGACGACT 121716

Query Match   5.3%;  Best Local Similarity  100.0%;   Score 20;  DB 2;  Length 122370;
Matches  20;  Conservative  0;  Mismatches  0;  Pred. No. 2;
              Indels  0;  Gaps  0;
```

RESULT 10

AP005251/c

```
LOCUS       AP005251       129591 bp    DNA    linear   HTG 24-MAY-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 8 clone
            clone:OSJNBb0011H15, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION   AP005251
VERSION     AP005251.1  GI:21165567
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group) DNA,
            clone:OSJNBb0011H15.
ORGANISM    Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
AUTHORS     Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE       Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, BAC
            clone:OSJNBb0011H15
JOURNAL     Published Only in Database (2002)
REFERENCE   2  (bases 1 to 129591)
AUTHORS     Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE       Direct Submission
JOURNAL     Submitted (23-MAY-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..129591
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="8"
                     /clone="OSJNBb0011H15"
BASE COUNT     36092 a  27555 c  28367 g  37453 t  124 others
ORIGIN

Qy          1 TACGGCTGCTTAAGACGACT 20
              ||||||||||||||||||||
Db     129591 TACGGCTGCTTAAGACGACT 16190

Query Match   5.3%;  Best Local Similarity  100.0%;   Score 20;  DB 2;  Length 129591;
Matches  20;  Conservative  0;  Mismatches  0;  Pred. No. 2;
              Indels  0;  Gaps  0;
```

RESULT 11

AL672277/c

```
LOCUS       AL672277       131682 bp    DNA    linear   PRI 20-AUG-2002
DEFINITION  Human DNA sequence from clone RP11-309M23 on chromosome X, complete
            sequence.
ACCESSION   AL672277
VERSION     AL672277.20  GI:22449770
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 131682)
AUTHORS     Clark,S.
TITLE       Direct Submission
JOURNAL     Submitted (20-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 22, 2002 this sequence version replaced gi:21911559.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
```

RESULT 12

Qy   172 AATAAAAACCTCCCTGCCAA 191
            |||||||||||||||||||||
Db 73926 AATAAAAACCTCCCTGCCAA 73907

Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Best Local Similarity 100.0%; Pred. No. 2;
Query Match 5.3%; Score 20; DB 9; Length 131682;

ORIGIN
BASE COUNT   34805 a 31378 c 30235 g 35264 t
VERSION   AC004520.2 GI:18855172
ACCESSION   AC004520
DEFINITION   Homo sapiens BAC clone CTB-119C2 from 7p15, complete sequence.
LOCUS   AC004520   132557 bp   DNA   linear   PRI 21-FEB-2002

KEYWORDS   HTG.
SOURCE   Homo sapiens.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 132557)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE   Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2 (bases 1 to 132557)
  AUTHORS   Dubuque,T., Smith,A., Elliott,G. and Harmon,G.
  TITLE   The sequence of Homo sapiens BAC clone CTB-119C2
  JOURNAL   Unpublished (2001)
REFERENCE   3 (bases 1 to 132557)
  AUTHORS   Waterston,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (01-APR-1998) Department of Genetics, Washington
             University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4 (bases 1 to 132557)
  AUTHORS   Waterston,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Department of Genetics, Washington
             University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5 (bases 1 to 132557)
  AUTHORS   Waterston,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (21-FEB-2002) Department of Genetics, Washington
             University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT   On Feb 21, 2002 this sequence version replaced gi:3004572.

             --------------- Genome Center
             Center: Washington University Genome Sequencing Center
             Center code: WUGSC
             Web site: http://genome.wustl.edu/gsc
             Contact: sapiens@watson.wustl.edu
             --------------- Summary Statistics
             Center project name: H_RG119C02

             NOTICE: This sequence may not represent the entire insert of this
             clone. It may be shorter because we only sequence overlapping
             clone sections once, or longer because we provide a small overlap
             between neighboring data submissions.

             This sequence was finished as follows unless otherwise noted:
             all regions were double stranded, sequenced with an alternate
             chemistry, or covered by high quality data (i.e, phred quality >=
             30); an attempt was made to resolve all sequencing problems, such
             as compressions and repeats; all regions were covered by sequence
             from more than one subclone; and the assembly was confirmed by
             restriction digest.

             MAPPING INFORMATION:
             The sequence of this clone was established as part of a mapping and
             sequencing collaboration between the NHGRI Chromosome 7 Mapping
             Project (Eric D. Green, Director), John D. McPherson in the
             Department of Genetics (Washington University), and the Washington
             University Genome Sequencing Center. For additional information
             about the map position of this sequence, see
             http://www.nhgri.nih.gov/DIR/GTB/CHR7/ send
             mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

             SOURCE INFORMATION:
             Clone CTB-119C2 is from the first release of the human BAC library
             CITB-978SK-B. The library contains cloned DNA from the male
             fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad.
             Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8
             (1996). This clone is available from Research Genetics, Inc.
             (http://www.resgen.com).
             VECTOR: pBeloBAC11
             Selection: chloramphenicol

             NEIGHBORING SEQUENCE INFORMATION:
             Actual start of this clone is at base position 1 of CTB-119C2;
             actual end is at base position 132572 of CTB-119C2.

FEATURES   Location/Qualifiers
     source   1..132557
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7p15"
                     /clone="CTB-119C2"
                     /clone_lib="CITB-978SK-B"
     repeat_region   43..351
                     /rpt_family="Alu"
     repeat_region   465..554
                     /rpt_family="L2"
     repeat_region   571..607
                     /rpt_family="Mariner"
     repeat_region   2067..2202
                     /rpt_family="L2"
     repeat_region   2203..2503
                     /rpt_family="Alu"
     repeat_region   2504..2622
                     /rpt_family="L2"
     repeat_region   2792..2967
                     /rpt_family="L2"
     repeat_region   3224..3454
                     /rpt_family="MIR"
     repeat_region   3975..4300
                     /rpt_family="Alu"
     repeat_region   4463..4765

Qy   172 AATAAAAACCTCCCTGCCAA 191

(RESULT 12 / AC004520 / c)

Db 73926

ORIGIN
     Chemistry or covered by high quality data (i.e, phred quality >=
     30); an attempt was made to resolve all sequencing problems, such
     as compressions and repeats; all regions were covered by at least
     one plasmid subclone or more than one M13 subclone; and the
     assembly was confirmed by restriction digest. The following
     abbreviations are used to associate primary accession numbers given
     in the feature table with their source databases: Em:, EMBL; Sw:,
     SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
     database can be found at
     http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
     was generated from part of bacterial clone contigs of human
     chromosome X, constructed by the Sanger Centre Chromosome X Mapping
     Group. Further information can be found at
     http://www.sanger.ac.uk/HGP/ChrX
     RP11-309M23 is from the library RPCI-11.2 constructed by the group
     of Pieter de Jong. For further details see
     http://www.chori.org/bacpac/home.htm

VECTOR: pBACe3.6.

FEATURES   Location/Qualifiers
     source   1..131682
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-309M23"
                     /clone_lib="RPCI-11.2"

```
repeat_region   4908..5230      /rpt_family="Alu"
repeat_region   5323..5630      /rpt_family="ERVL"
repeat_region   6328..6422      /rpt_family="Alu"
repeat_region   6937..7144      /rpt_family="L2"
repeat_region   7409..7490      /rpt_family="MIR"
repeat_region   7708..7953      /rpt_family="L2"
repeat_region   7954..8040      /rpt_family="Alu"
repeat_region   8272..8569      /rpt_family="Alu"
repeat_region   8622..8696      /rpt_family="MIR"
repeat_region   8708..9164      /rpt_family="MIR"
repeat_region   9180..9274      /rpt_family="ERVL"
repeat_region   11518..11717    /rpt_family="MIR"
repeat_region   11718..11817    /rpt_family="MER1_type"
repeat_region   13943..14241    /rpt_family="Alu"
repeat_region   14463..14687    /rpt_family="Alu"
repeat_region   15533..15784    /rpt_family="Alu"
repeat_region   15785..15908    /rpt_family="L1"
repeat_region   15909..15986    /rpt_family="Alu"
repeat_region   16143..16300    /rpt_family="L1"
repeat_region   16480..16528    /rpt_family="L1"
repeat_region   16706..16987    /rpt_family="MIR"
repeat_region   17781..18181    /rpt_family="Alu"
repeat_region   18493..18796    /rpt_family="Alu"
repeat_region   19290..19470    /rpt_family="Alu"
repeat_region   20005..20824    /rpt_family="MIR"
repeat_region   20825..21126    /rpt_family="ERV1"
repeat_region   21127..21529    /rpt_family="Alu"
repeat_region   22556..22787    /rpt_family="ERV1"
repeat_region   22795..22839    /rpt_family="MER1_type"
repeat_region   22860..23070    /rpt_family="MER1_type"
repeat_region   23851..24034    /rpt_family="MIR"
repeat_region   24038..24353    /rpt_family="Alu"
repeat_region   25234..25546    /rpt_family="Alu"
repeat_region   26003..26451    /rpt_family="Alu"
repeat_region   26626..26914    /rpt_family="ERVL"
repeat_region   27468..27532    /rpt_family="Alu"
repeat_region                   /rpt_family="MER1_type"

repeat_region   28341..28410    /rpt_family="L2"
repeat_region   28546..28691    /rpt_family="MIR"
repeat_region   29423..29500    /rpt_family="MIR"
repeat_region   29635..29987    /rpt_family="MaLR"
repeat_region   30261..30406    /rpt_family="Alu"
repeat_region   30407..30601    /rpt_family="Alu"
repeat_region   30602..30765    /rpt_family="Alu"
repeat_region   31508..31716    /rpt_family="Alu"
repeat_region   33321..33438    /rpt_family="L1"
repeat_region   33600..33809    /rpt_family="MIR"
repeat_region                   /rpt_family="L1"
```

```
Qy   84  CACAAGCTTGAATGTGTGTT 103
Db 13715 CACAAGCTTGAATGTGTGTT 13696

Query Match                5.3%;   Score 20;  DB 9;   Length 132557;
Best Local Similarity   100.0%;    Pred. No. 2;
Matches  20;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
RESULT 13
AC046149/c
LOCUS       AC046149            156469 bp    DNA     linear   HTG 16-OCT-2001
DEFINITION  Mus musculus chromosome 10 clone RP23-310B11, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC046149
VERSION     AC046149.5  GI:16118092
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 156469)
  AUTHORS   Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
            Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
            Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
            Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
            Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
            Fermandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
            Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
            Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
            Kovar,C., Liu,J., Liu,W., Louiseged,H., Lozado,R.J., Martin,R.,
            Massey,E., McLeod,M.P., Nguyen,J., Mei,G., Moore,S., Morgan,M.,
            Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
            Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
            Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
            Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
            Worley,K.C. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156469)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 14, 2001 this sequence version replaced gi:11094642.
            --------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            --------- Project Information
```

```
Center project name: MAEF
Center clone name: RP23-310B11
------- Summary Statistics -------
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 14589 bases at least Q40
Consensus quality: 154088 bases at least Q30
Consensus quality: 157075 bases at least Q20
Consensus quality: 159414; sum-of-contigs estimation
Estimated insert size: 159414; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 3.9x in Q20 bases; sum-of-contigs estimation
---------------------------------
*  NOTE: Estimated insert size may differ from sequence length
*    (see http://www.hgsc.bcm.tmc.edu/docs/GenBank_draft_data.html).
*  NOTE: This is a 'working draft' sequence. It currently
*  consists of 13 contigs. The true order of the pieces
*  is not known and their order in this sequence record is
*  arbitrary. Gaps between the contigs are represented as
*  runs of N, but the exact sizes of the gaps are unknown.
*  This record will be updated with the finished sequence
*  as soon as it is available and the accession number will
*  be preserved.
*
*          1         38620: contig of 38620 bp in length
*      38621         38720: gap of unknown length
*      38721         62174: contig of 23454 bp in length
*      62175         62274: gap of unknown length
*      62275         83257: contig of 20983 bp in length
*      83258         83357: gap of unknown length
*      83358         96441: contig of 13084 bp in length
*      96442         96541: gap of unknown length
*      96542        108526: contig of 11985 bp in length
*     108527        108626: gap of unknown length
*     108627        119782: contig of 11156 bp in length
*     119783        119882: gap of unknown length
*     119883        129550: contig of 9668 bp in length
*     129551        129650: gap of unknown length
*     129651        136363: contig of 6713 bp in length
*     136364        136463: gap of unknown length
*     136464        142311: contig of 5848 bp in length
*     142312        142411: gap of unknown length
*     142412        147231: contig of 4820 bp in length
*     147232        147331: gap of unknown length
*     147332        151809: contig of 4478 bp in length
*     151810        151909: gap of unknown length
*     151910        154013: contig of 2104 bp in length
*     154014        154113: gap of unknown length
*     154114        156469: contig of 2356 bp in length.

FEATURES             Location/Qualifiers
     source          1..156469
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="10"
                     /clone="RP23-310B11"

BASE COUNT    39005 a  38403 c  38896 g  38948 t  1217 others
ORIGIN
```

```
Qy    344 TCCCGGGAAGGAAGGTACCA 363
              |||||||||||||||||||
Db 144703 TCCCGGGAAGGAAGGTACCA 144684
```

```
RESULT 14
AC009295/c
LOCUS               159108
DEFINITION    Mus musculus chromosome 10, clone RP21-39C4, complete sequence.
ACCESSION    AC009295
VERSION        AC009295.14  GI:9280781
KEYWORDS    HTG.
```

```
Query Match      5.3%;   Score 20;   DB 2;   Length 156469;
Best Local Similarity   100.0%;   Pred. No. 2;
Matches 20;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;
```

```
LOCUS       AC009295    159108 bp    DNA    linear    ROD 01-AUG-2000
```

```
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 159108)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 10, clone RP21-39C4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 159108)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,K., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Nhloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 159108)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Illev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,M., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,K., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,D., Tirrell,A., Travers,M., Trigillo,J.,
            Vassliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 19, 2000 this sequence version replaced gi:8571747.
            All repeats were identified using RepeatMasker.
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            Smit, A.F.A. & Green, P. (1996-1997)
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: L760
            Center clone name: 39_C4
            ---------------

FEATURES             Location/Qualifiers
     source          1..159108
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="10"
                     /clone="RP21-39C4"
                     /map="10"
                     /clone_lib="RPCI-21 Female Mouse PAC"
```

```
repeat_region   2048..2155
                /rpt_family="GC_rich"
repeat_region   2228..2290
                /rpt_family="GC_rich"
repeat_region   complement(7061..7239)
                /rpt_family="AluJo"
repeat_region   complement(7448..7557)
                /rpt_family="FLAM_A"
repeat_region   7581..7618
                /rpt_family="FLAM_A"
repeat_region   11048..11067
                /rpt_family="(CAAAA)n"
repeat_region   11346..11393
                /rpt_family="(GGGA)n"
repeat_region   11783..12816
                /rpt_family="(CA)n"
repeat_region   complement(12904..12963)
                /rpt_family="(TTTA)n"
repeat_region   14619..14709
                /rpt_family="MIR"
repeat_region   15382..15475
                /rpt_family="FLAM_C"
repeat_region   complement(15532..15625)
                /rpt_family="(TTCTC)n"
repeat_region   16802..16860
                /rpt_family="AluJb"
repeat_region   17130..17310
                /rpt_family="AluSp"
repeat_region   20035..20128
                /rpt_family="(CATA)n"
repeat_region   20387..20435
                /rpt_family="FLAM_C"
repeat_region   20436..20528
                /rpt_family="(CA)n"
repeat_region   20837..20867
                /rpt_family="(GA)n"
repeat_region   21617..21641
                /rpt_family="(TGG)n"
repeat_region   22773..22821
                /rpt_family="AT_rich"
repeat_region   22878..23043
                /rpt_family="(TTTTA)n"
repeat_region   23211..23273
                /rpt_family="FRAM"
repeat_region   23448..23475
                /rpt_family="CT_rich"
repeat_region   24116..24201
                /rpt_family="(TG)n"
repeat_region   24359..24450
                /rpt_family="AluJo"
repeat_region   25402..25429
                /rpt_family="L1MB2"
repeat_region   25429..25494
                /rpt_family="(TG)n"
repeat_region   26461..26508
                /rpt_family="AluS"
repeat_region   27167..27197
                /rpt_family="(GA)n"
repeat_region   27210..27269
                /rpt_family="(CA)n"
repeat_region   27633..27673
                /rpt_family="AT_rich"
repeat_region   28793..28877
                /rpt_family="MIR"
repeat_region   28921..28958
                /rpt_family="AluS"
repeat_region   complement(29020..29094)
                /rpt_family="AT_rich"
repeat_region   31216..31256
                /rpt_family="MIR"
repeat_region   complement(31814..32710)
                /rpt_family="CT_rich"
repeat_region   33264..33314
                /rpt_family="L1MA5A"

repeat_region   35155..35191
                /rpt_family="HERV15I|HERV15I"
repeat_region   35254..35300
                /rpt_family="AT_rich"
repeat_region   35725..35764
                /rpt_family="(GGGGA)n"
repeat_region   38626..38803
                /rpt_family="(TG)n"
repeat_region   39467..39493
                /rpt_family="(CA)n"
repeat_region   40008..40048
                /rpt_family="AT_rich"
repeat_region   complement(40052..40170)
                /rpt_family="(TTTC)n"
repeat_region   40298..40417
                /rpt_family="5S"
repeat_region   complement(42703..42930)
                /rpt_family="(GGAA)n"
repeat_region   43475..43511
                /rpt_family="L1MA6"
repeat_region   43513..43596
                /rpt_family="(TCTG)n"
repeat_region   43596..43672
                /rpt_family="(TG)n"
repeat_region   complement(44004..44081)
                /rpt_family="(GA)n"
repeat_region   48910..48930
                /rpt_family="L1MA5A"
repeat_region   complement(49480..49658)
                /rpt_family="(TG)n"
repeat_region   51864..51890
                /rpt_family="AluJo"
repeat_region   52072..52212
                /rpt_family="AT_rich"
repeat_region   52299..52326
                /rpt_family="FRAM"
repeat_region   52414..52521
                /rpt_family="(TTCTCC)n"
repeat_region   52664..52950
                /rpt_family="AluS"
repeat_region   52951..52980
                /rpt_family="AluJo"
repeat_region   52981..52995
                /rpt_family="(CCAA)n"
repeat_region   complement(54449..54619)
                /rpt_family="AluJo"
repeat_region   58720..58753
                /rpt_family="AluJo"
                /rpt_family="(TG)n"
```

```
REFERENCE    1  (bases 1 to 162394)
             Rattus.
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM     Rattus norvegicus
SOURCE       Norway rat.
             Rattus norvegicus
KEYWORDS     HTG; HTGS_PHASE1.
VERSION      AC119775.4  GI:21747184
ACCESSION    AC119775
             ***, 61 unordered pieces.
DEFINITION   Rattus norvegicus clone CH230-464J18, *** SEQUENCING IN PROGRESS
LOCUS        AC119775             162394 bp    DNA     linear   HTG 18-JUL-2002
AC119775/c
RESULT 15

Db 105501  TCCCGGGAAGGAAGGTACCA  105482
           ||||||||||||||||||||
Qy    344  TCCCGGGAAGGAAGGTACCA  363

Matches   20;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
Best Local Similarity  100.0%;  Pred. No. 2;
Query Match  5.3%;  Score 20;  DB 10;  Length 159108;
```

AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
          Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
          Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
          Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
          Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
          Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
          Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
          Cleveland,C.D., Cox,C.H., Coyle,M.D., Dathorne,S.R., David,R.,
          Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
          Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
          Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
          Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
          Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
          Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
          Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
          Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
          Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
          Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
          Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
          Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
          Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L., Lewis,L.,
          Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
          Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
          Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
          Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
          Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
          Moser,M., Neal,D., Newson,J., Newton,N., Nguyen,A., Nguyen,N.,
          Nguyen,N., Nickerson,E., Nkokenkwo,S., Oguh,M., Okwuonu,G.,
          Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
          Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
          Rives,J., Rojas,A., Rojubokan,I., Rolfe,H., Ruiz,S.J., Savery,G.,
          Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
          Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
          Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
          Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
          Usman,I.K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
          Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
          Williams,G., Williamson,A., Wieczyk,R., Wooden,S., Worley,K.,
          Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
          Weinstock,G. and Gibbs,R.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 162394)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (02-MAY-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 162394)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
          On Jul 14, 2002 this sequence version replaced gi:20467838.
COMMENT   Genome Center
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
          -----------------Project Information
          Center project name: GVQM
          Center clone name: CH230-464J18
          -----------------Summary Statistics
          Sequencing vector: plasmid;
          Chemistry: Dye-terminator Big Dye: 100% of reads
          Assembly program: Phrap; version 0.990329
          Consensus quality: 104933 bases at least Q40
          Consensus quality: 112584 bases at least Q30
          Consensus quality: 117408 bases at least Q20

          ----
          * NOTE: Estimated insert size may differ from sequence length
          * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
        1     1004: contig of 1004 bp in length
     1005     1104: gap of unknown length
     1105     2121: contig of 1017 bp in length
     2122     2221: gap of unknown length
     2222     3241: contig of 1020 bp in length
     3242     3341: gap of unknown length
     3342     4689: contig of 1348 bp in length
     4690     4789: gap of unknown length
     4790     6424: contig of 1635 bp in length
     6425     6524: gap of unknown length
     6525     8023: contig of 1499 bp in length
     8024     8123: gap of unknown length
     8124     9618: contig of 1495 bp in length
     9619     9718: gap of unknown length
     9719    10759: contig of 1041 bp in length
    10760    10859: gap of unknown length
    10860    12638: contig of 1779 bp in length
    12639    12738: gap of unknown length
    12739    14083: contig of 1345 bp in length
    14084    14183: gap of unknown length
    14184    15891: contig of 1708 bp in length
    15892    15991: gap of unknown length
    15992    17623: contig of 1632 bp in length
    17624    17723: gap of unknown length
    17724    18962: contig of 1239 bp in length
    18963    19062: gap of unknown length
    19063    20849: contig of 1787 bp in length
    20850    20949: gap of unknown length
    20950    22169: contig of 1220 bp in length
    22170    22269: gap of unknown length
    22270    23483: contig of 1214 bp in length
    23484    23583: gap of unknown length
    23584    24875: contig of 1292 bp in length
    24876    24975: gap of unknown length
    24976    26670: contig of 1695 bp in length
    26671    26770: gap of unknown length
    26771    27914: contig of 1144 bp in length
    27915    28014: gap of unknown length
    28015    29124: contig of 1110 bp in length
    29125    29224: gap of unknown length
    29225    31386: contig of 2162 bp in length
    31387    31486: gap of unknown length
    31487    32500: contig of 1014 bp in length
    32501    32600: gap of unknown length
    32601    34344: contig of 1745 bp in length
    34345    34445: gap of unknown length
    34446    36597: contig of 2152 bp in length
    36598    36697: gap of unknown length
    36698    38768: contig of 2071 bp in length
    38769    38868: gap of unknown length
    38869    40546: contig of 1678 bp in length
    40547    40646: gap of unknown length
    40647    41770: contig of 1124 bp in length
    41771    41870: gap of unknown length
    41871    43735: contig of 1865 bp in length
    43736    43835: gap of unknown length
    43836    45206: contig of 1371 bp in length
    45207    45306: gap of unknown length
    45307    47814: contig of 2508 bp in length
    47815    47914: gap of unknown length
    47915    49406: contig of 1492 bp in length
    49407    49506: gap of unknown length
    49507    51910: contig of 2404 bp in length
    51911    52010: gap of unknown length
    52011    54105: contig of 2095 bp in length
```

```
Query Match                     5.3%;  Score 20;  DB 2;  Length 162394;
Best Local Similarity  100.0%;  Pred. No. 2;
Matches  20;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    210 CTCAACAATGCTTTGACTGTT 229
          |||||||||||||||||||||
Db  79021 CTCAACAATGCTTTGACTGTT 79002
```

```
*    54106    54205: gap of unknown length
*    54206    56954: contig of 2749 bp in length
*    56955    57054: gap of unknown length
*    57055    59393: contig of 2339 bp in length
*    59394    59493: gap of unknown length
*    59494    61754: contig of 2261 bp in length
*    61755    61854: gap of unknown length
*    61855    64055: contig of 2201 bp in length
*    64056    64155: gap of unknown length
*    64156    67129: contig of 2974 bp in length
*    67130    67229: gap of unknown length
*    67230    69622: contig of 2393 bp in length
*    69623    69722: gap of unknown length
*    69723    72435: contig of 2713 bp in length
*    72436    72535: gap of unknown length
*    72536    75742: contig of 3207 bp in length
*    75743    75842: gap of unknown length
*    75843    77891: contig of 2049 bp in length
*    77892    77991: gap of unknown length
*    77992    80486: contig of 2495 bp in length
*    80487    80586: gap of unknown length
*    80587    83986: contig of 3400 bp in length
*    83987    84086: gap of unknown length
*    84087    86359: contig of 2273 bp in length
*    86360    86459: gap of unknown length
*    86460    91249: contig of 4790 bp in length
*    91250    91349: gap of unknown length
*    91350    94806: contig of 3457 bp in length
*    94807    94906: gap of unknown length
*    94907    98994: contig of 4088 bp in length
*    98995    99094: gap of unknown length
*    99095   105047: contig of 5953 bp in length
*   105048   105147: gap of unknown length
*   105148   108668: contig of 3521 bp in length
*   108669   108768: gap of unknown length
*   108769   111829: contig of 3061 bp in length
*   111830   111929: gap of unknown length
*   111930   116397: contig of 4468 bp in length
*   116398   116497: gap of unknown length
*   116498   122101: contig of 5604 bp in length
*   122102   122201: gap of unknown length
*   122202   126352: contig of 4151 bp in length
```

Search completed: January 31, 2003, 04:46:58
Job time : 3388 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: January 31, 2003, 01:07:09

Searched: 2475311 seqs, 567426730 residues

Scoring table: Gapop 10.0 , Gapext 1.0
Scoring table: IDENTITY_NUC
Sequence: 1 tacggctgcttaagacgact.........taccattgagagaccatgcg 378
Perfect score: 378
Title:
Sequence: US-09-803-719-222

Database : Pending_Patents_NA_New:*

```
Listing first 45 summaries
Maximum Match 100%
1:  /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
2:  /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3:  /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4:  /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5:  /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6:  /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
7:  /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*
```

Post-processing: Minimum Match 0%

Maximum DB seq length: 2000000000
Minimum DB seq length: 0

Total number of hits satisfying chosen parameters: 4950622

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	35.2	9.3	1671	1	PCT-US02-10366-23	Sequence 23, Appl
	2	35.2	9.3	1677	1	PCT-US02-10366-21	Sequence 21, Appl
	3	35.2	9.3	2067	1	PCT-US02-10366-27	Sequence 27, Appl
	4	35.2	9.3	2130	1	PCT-US02-10366-25	Sequence 25, Appl
	5	33.6	8.9	492	6	US-10-276-774-299	Sequence 299, App
	6	33.6	8.9	1629	1	PCT-US02-38526-589	Sequence 589, App
	7	33.4	8.9	1640	1	PCT-US02-38526-587	Sequence 587, App
	8	33.4	8.9	158	1	US-09-513-999C-33882	Sequence 33882, A
c	9	32.6	8.6	45450	1	PCT-US02-33700-7	Sequence 7, Appli
	10	32.6	8.6	45450	1	PCT-US02-33700-7	Sequence 7, Appli
c	11	32.4	8.6	334	5	US-10-271-416-7	Sequence 7, Appli
c	12	32.2	8.5	264	5	US-09-531-113-5935	Sequence 5935, App
c	13	32	8.5	410	5	US-09-950-084-952	Sequence 952, App
	14	31.6	8.4	2966	6	US-09-513-999C-30304	Sequence 30304, A
	15	31.6	8.4	3011	6	US-10-294-443-103	Sequence 103, App
	16	31.6	8.4	3701	6	US-10-294-443-10	Sequence 10, Appl
c	17	30.8	8.1	1083	6	US-10-294-443-112	Sequence 112, App
c	18	30.8	8.1	1083	5	US-09-724-676-6004	Sequence 6004, Ap
c	19	30.8	8.1	1180	5	US-09-724-676A-6004	Sequence 6004, Ap
c	20	30.8	8.1	1180	6	US-09-724-676-6035	Sequence 6035, Ap
c	21	30.8	8.1	1208	6	US-10-276-774-1101	Sequence 1101, Ap
c	22	30.8	8.1	1392	5	US-09-724-676-5990	Sequence 5990, Ap
c	23	30.8	8.1	1392	5	US-09-724-676A-5990	Sequence 5990, Ap
c	24	30.8	8.1	1489	5	US-09-724-676-6021	Sequence 6021, Ap
c	25	30.8	8.1	1489	5	US-09-724-676A-6021	Sequence 6021, Ap
c	26	30.8	8.1	1770	5	US-09-724-676-5980	Sequence 5980, Ap
c	27	30.8	8.1	1770	5	US-09-724-676A-5980	Sequence 5980, Ap
c	28	30.8	8.1	1867	5	US-09-724-676-6011	Sequence 6011, Ap
c	29	30.8	8.1	1867	5	US-09-724-676A-6011	Sequence 6011, Ap
c	30	30.8	8.1	2347	5	US-10-240-425-389	Sequence 389, App
c	31	30.4	8.0	2771	5	US-09-724-676-37523	Sequence 37523, A
c	32	30.4	8.0	2771	5	US-09-724-676A-37523	Sequence 37523, A
c	33	30.4	8.0	2866	5	US-09-724-676-37524	Sequence 37524, A
c	34	30.4	8.0	2866	5	US-09-724-676A-37524	Sequence 37524, A
c	35	30.2	8.0	36303	6	US-10-152-724-24	Sequence 24, Appl
c	36	30.2	8.0	36303	1	PCT-US02-32700-8	Sequence 8, Appli
c	37	30.2	8.0	302603	6	US-10-152-724A-24	Sequence 24, Appl
c	38	30.2	7.9	302603	1	US-10-271-416-8	Sequence 8, Appli
c	39	30	7.9	14759	6	US-10-346-723-1	Sequence 1, Appli
c	40	30	7.9	26159	6	US-10-144-779-131	Sequence 131, App
c	41	29.8	7.9	75829	6	US-09-531-113-31549	Sequence 31549, A
c	42	29.8	7.9	472	6	US-10-144-779-230	Sequence 230, App
c	43	29.8	7.9	1457	6	US-09-289-757-41	Sequence 41, Appl
c	44	29.8	7.9	859705	5	US-09-948-124-60	Sequence 60, Appl
c	45	29.6	7.8	648	5	US-09-513-999C-3839	Sequence 3839, Ap

ALIGNMENTS

RESULT 1
PCT-US02-10366-23
Sequence 23, Application PCT/US0210366
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glennda
; APPLICANT: Guo, Xiaojla
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Tcherney, Velizar
; APPLICANT: Gangolli, Esha
; APPLICANT: Vernet, Corine
; APPLICANT: Spytek, Kimberly
; APPLICANT: Malyankar, Uriel
; APPLICANT: Patturajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; APPLICANT: Heyes, Melvyn
; APPLICANT: Ju, Jingfang
; APPLICANT: Peyman, John
; APPLICANT: Catterton, Elina
; APPLICANT: MacDougall, John
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Mazur, Ann
; TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES,
; TITLE OF INVENTION: NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE ANTIGENS, AND METHODS OF USE
; FILE REFERENCE: 21402-322A-061
; CURRENT APPLICATION NUMBER: PCT/US02/10366
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 10/114153
; PRIOR FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: 60/281086
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/281906
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/282020
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/282930
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/283512

```
~ PRIOR FILING DATE: 2001-04-12
~ PRIOR APPLICATION NUMBER: 60/283444
~ PRIOR FILING DATE: 2001-04-12
~ PRIOR APPLICATION NUMBER: 60/283657
~ PRIOR FILING DATE: 2001-04-13
~ PRIOR APPLICATION NUMBER: 60/283710
~ PRIOR FILING DATE: 2001-04-13
~ PRIOR APPLICATION NUMBER: 60/283678
~ PRIOR FILING DATE: 2001-04-13
~ Remaining Prior Application data removed - See File Wrapper or PALM.
~ NUMBER OF SEQ ID NOS: 251
PCT-US02-10366-23
~ SEQ ID NO 23
~ LENGTH: 1671
~ TYPE: DNA
~ ORGANISM: Homo sapiens
~ FEATURE:
~ NAME/KEY: CDS
~ LOCATION: (31)..(1663)
PCT-US02-10366-23

Query Match            9.3%;  Score 35.2;  DB 1;  Length 1671;
Best Local Similarity   53.7%;  Pred. No. 0.34;
Matches  73;  Conservative  0;  Mismatches  63;  Indels  0;  Gaps  0;

Qy      91 TTGAATGTGTGTTCTGGGAGCTGAAGGATGCACGGTTGTTAAGCCCCTGTTCTTTTTCCGTT 150
            | || |  || |||  |||| || ||||  || ||| ||||| |||  ||| |||  ||
Db    1092 TGGAACCTCTACGCTCTTGGTTGTTGGAGTCATCAGATACCTGGGTTATTTCCAGGC 1151

Qy     151 GTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTATGC 210
            | |||| ||| ||  ||||| | |  |||| ||||  ||||||||| ||||| |||| |||
Db    1152 ATATAATGTACTGATTTTAACAATGCAGGCCCTCACTGCCAAAAAGTTCTTCGGTTTTGTGC 1211

Qy     211 TCAACATGCTTTGACT 226
            ||||| ||||| ||| |
Db    1212 TTGTCCTGGTATGATT 1227
```

```
RESULT 2
PCT-US02-10366-21
; Sequence 21, Application PC/US0210366
; GENERAL INFORMATION:
~ APPLICANT: Padigaru, Muralidhara
~ APPLICANT: Shenoy, Suresh
~ APPLICANT: Kekuda, Ramesh
~ APPLICANT: Rastelli, Luca
~ APPLICANT: Mezes, Peter
~ APPLICANT: Smithson, Glennda
~ APPLICANT: Guo, Xiaojia
~ APPLICANT: Gerlach, Valerie
~ APPLICANT: Casman, Stacie
~ APPLICANT: Boldog, Ferenc
~ APPLICANT: Li, Li
~ APPLICANT: Zerhusen, Bryan
~ APPLICANT: Tchernev, Velizar
~ APPLICANT: Gangolli, Esha
~ APPLICANT: Vernet, Corine
~ APPLICANT: Spytek, Kimberly
~ APPLICANT: Malyankar, Uriel
~ APPLICANT: Patturajan, Meera
~ APPLICANT: Miller, Charles
~ APPLICANT: Taupier, Raymond J. Jr.
~ APPLICANT: Heyes, Melvyn
~ APPLICANT: Ju, Jingfang
~ APPLICANT: Peyman, John
~ APPLICANT: Catterton, Ellna
~ APPLICANT: MacDougall, John
~ APPLICANT: Edinger, Shlomit
~ APPLICANT: Stone, David
~ APPLICANT: Mazur, Ann
~ TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES,
~ TITLE OF INVENTION: NUCLEIC
~ TITLE OF INVENTION: ACIDS ENCODING THE ANTIGENS, AND METHODS OF USE
~ FILE REFERENCE: 21102-322A-061
```

```
~ CURRENT APPLICATION NUMBER: PCT/US02/10366
~ CURRENT FILING DATE: 2002-04-03
~ PRIOR APPLICATION NUMBER: 10/114153
~ PRIOR FILING DATE: 2002-04-02
~ PRIOR APPLICATION NUMBER: 60/281086
~ PRIOR FILING DATE: 2001-04-03
~ PRIOR APPLICATION NUMBER: 60/281906
~ PRIOR FILING DATE: 2001-04-05
~ PRIOR APPLICATION NUMBER: 60/282020
~ PRIOR FILING DATE: 2001-04-06
~ PRIOR APPLICATION NUMBER: 60/282930
~ PRIOR FILING DATE: 2001-04-10
~ PRIOR APPLICATION NUMBER: 60/283512
~ PRIOR FILING DATE: 2001-04-12
~ PRIOR APPLICATION NUMBER: 60/283444
~ PRIOR FILING DATE: 2001-04-12
~ PRIOR APPLICATION NUMBER: 60/283657
~ PRIOR FILING DATE: 2001-04-13
~ PRIOR APPLICATION NUMBER: 60/283710
~ PRIOR FILING DATE: 2001-04-13
~ PRIOR APPLICATION NUMBER: 60/283678
~ PRIOR FILING DATE: 2001-04-13
~ Remaining Prior Application data removed - See File Wrapper or PALM.
~ NUMBER OF SEQ ID NOS: 251
SEQ ID NO 21
~ LENGTH: 1677
~ TYPE: DNA
~ ORGANISM: Homo sapiens
~ FEATURE:
~ NAME/KEY: CDS
~ LOCATION: (25)..(1645)
PCT-US02-10366-21

Query Match            9.3%;  Score 35.2;  DB 1;  Length 1677;
Best Local Similarity   53.7%;  Pred. No. 0.34;
Matches  73;  Conservative  0;  Mismatches  63;  Indels  0;  Gaps  0;

Qy      91 TTGAATGTGTGTTCTGGGAGCTGAAGGATGCACGGTTGTTAAGCCCCTGTTCTTTTTCCGTT 150
            | || |  || |||  |||| || ||||  || ||| ||||| |||  ||| |||  ||
Db    1074 TGGAACCTCTACGCTCTTGGTTGTTGGAGTCATCAGATACCTGGGTTATTTCCAGGC 1133

Qy     151 GTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTATGC 210
            | |||| ||| ||  ||||| | |  |||| ||||  ||||||||| ||||| |||| |||
Db    1134 ATATAATGTACTGATTTTAACAATGCAGGCCCTCACTGCCAAAAAGTTCTTCGGTTTTGTGC 1193

Qy     211 TCAACATGCTTTGACT 226
            ||||| ||||| ||| |
Db    1194 TTGTCCTGGTATGATT 1209
```

```
RESULT 3
PCT-US02-10366-27
; Sequence 27, Application PC/US0210366
; GENERAL INFORMATION:
~ APPLICANT: Padigaru, Muralidhara
~ APPLICANT: Shenoy, Suresh
~ APPLICANT: Kekuda, Ramesh
~ APPLICANT: Rastelli, Luca
~ APPLICANT: Mezes, Peter
~ APPLICANT: Smithson, Glennda
~ APPLICANT: Guo, Xiaojia
~ APPLICANT: Gerlach, Valerie
~ APPLICANT: Casman, Stacie
~ APPLICANT: Boldog, Ferenc
~ APPLICANT: Li, Li
~ APPLICANT: Zerhusen, Bryan
~ APPLICANT: Tchernev, Velizar
~ APPLICANT: Gangolli, Esha
~ APPLICANT: Vernet, Corine
~ APPLICANT: Spytek, Kimberly
~ APPLICANT: Malyankar, Uriel
~ APPLICANT: Patturajan, Meera
~ APPLICANT: Miller, Charles
```

APPLICANT: Taupier, Raymond J. Jr.
; APPLICANT: Heyes, Melvyn
; APPLICANT: Ju, Jingfang
; APPLICANT: Peyman, John
; APPLICANT: Catterton, Elina
; APPLICANT: MacDougall, John
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Mazur, Ann
TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES,
TITLE OF INVENTION: NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE ANTIGENS, AND METHODS OF USE
FILE REFERENCE: 21402-322A-061
CURRENT APPLICATION NUMBER: PCT/US02/10366
CURRENT FILING DATE: 2002-04-03
PRIOR APPLICATION NUMBER: 10/114153
PRIOR FILING DATE: 2002-04-02
PRIOR APPLICATION NUMBER: 60/281086
PRIOR FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: 60/281906
PRIOR FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: 60/282020
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: 60/282930
PRIOR FILING DATE: 2001-04-10
PRIOR APPLICATION NUMBER: 60/283512
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 60/283444
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 60/283657
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/283710
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/283678
PRIOR FILING DATE: 2001-04-13
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 251

SEQ ID NO 27
LENGTH: 2067
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (93)..(1791)

Query Match   9.3%;   Score 35.2;   DB 1;   Length 2067;
Best Local Similarity   53.7%;   Pred. No. 0.37;
Matches   73;   Conservative   0;   Mismatches   63;   Indels   0;   Gaps   0;

```
Qy    91 TTGAATGTGTGTTCTGGGAGCTGAAGGATGCACGGTTGTTAAGCCCCTGTTTCTTTTCCGTT 150
          | ||||  ||||  || ||||| || ||| ||  | || | ||| ||||||| || | ||
Db  1238 TGGAACCTCTACGCTCTTGGTTTGGTTGGAGTCATCAGATACCTGGGTTATTCCAGGC 1297

Qy   151 GTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTTATGC 210
          ||||| ||| ||| || |||  ||||||||||||||| ||||||||||| ||||||| |||
Db  1298 ATATAATGTACTGATTTTAACAATGCAGGCCTCACTGCCAAAAGTTCTTCGGTTTTGTGC 1357

Qy   211 TCAACATGCTTTGACT 226
          || | |||  ||| |
Db  1358 TTGTGCTGGTATGATT 1373
```

GENERAL INFORMATION:
APPLICANT: Padigaru, Muralidhara
APPLICANT: Shenoy, Suresh
APPLICANT: Kekuda, Ramesh
APPLICANT: Rastelli, Luca
APPLICANT: Mezes, Peter
APPLICANT: Smithson, Glennda

RESULT 4
PCT-US02-10366-25
Sequence 25, Application PCT/US0210366

APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Tcherney, Velizar
; APPLICANT: Gangolli, Esha
; APPLICANT: Vernet, Corine
; APPLICANT: Spytek, Kimberly
; APPLICANT: Malyankar, Uriel
; APPLICANT: Pattutajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; APPLICANT: Heyes, Melvyn
; APPLICANT: Ju, Jingfang
; APPLICANT: Peyman, John
; APPLICANT: Catterton, Elina
; APPLICANT: MacDougall, John
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Mazur, Ann
TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES,
TITLE OF INVENTION: NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE ANTIGENS, AND METHODS OF USE
FILE REFERENCE: 21402-322A-061
CURRENT APPLICATION NUMBER: PCT/US02/10366
CURRENT FILING DATE: 2002-04-03
PRIOR APPLICATION NUMBER: 10/114153
PRIOR FILING DATE: 2002-04-02
PRIOR APPLICATION NUMBER: 60/281086
PRIOR FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: 60/281906
PRIOR FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: 60/282020
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: 60/282930
PRIOR FILING DATE: 2001-04-10
PRIOR APPLICATION NUMBER: 60/283512
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 60/283444
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 60/283657
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/283710
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/283678
PRIOR FILING DATE: 2001-04-13
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 251

SEQ ID NO 25
LENGTH: 2130
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (31)..(1645)

Query Match   9.3%;   Score 35.2;   DB 1;   Length 2130;
Best Local Similarity   53.7%;   Pred. No. 0.37;
Matches   73;   Conservative   0;   Mismatches   63;   Indels   0;   Gaps   0;

```
Qy    91 TTGAATGTGTGTTCTGGGAGCTGAAGGATGCACGGTTGTTAAGCCCCTGTTTCTTTTCCGTT 150
          | ||||  ||||  || ||||| || ||| ||  | || | ||| ||||||| || | ||
Db  1092 TGGAACCTCTACGCTCTTGGTTTGGTTGGAGTCATCAGATACCTGGGTTATTCCAGGC 1151

Qy   151 GTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTTATGC 210
          ||||| ||| ||| || |||  ||||||||||||||| ||||||||||| ||||||| |||
Db  1152 ATATAATGTACTGATTTTAACAATGCAGGCCTCACTGCCAAAAGTTCTTCGGTTTTGTGC 1211

Qy   211 TCAACATGCTTTGACT 226
          || | |||  ||| |
```

Db   1212 TTGTGCTGGTATGATT 1227

RESULT 5
US-10-276-774-299
Sequence 299, Application US/10276774
GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y. Tom et al
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10,276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 299
; LENGTH: 492
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-276-774-299

Query Match   8.9%; Score 33.6; DB 6; Length 492;
Best Local Similarity   52.9%; Pred. No. 0.75;
Matches   72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy   91 TTGAATGTGTTCTGGGACTGAAGGATGCACGGTTGTTGTTAAGCCCTGTTCTTTTCCGTT 150
          ||| ||| |||||||| |||| ||| || |||||||||||| ||| ||||| || |||
Db   51 TGGAACCCTCTACGCTCTTCGGTTTGGTGGAGTCATCAGATACCTGGGTTATTTCCAGGC 110

Qy   151 GTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTATGC 210
          ||||||| |||| || |||||||| | ||||| |||||||||||||| ||||||||||||
Db   111 ATATAATGTGCTGATTTAACAATGCAGCGCCTCACTGCCAAAAAGTTCTTCGGTTTTGTGC 170

Qy   211 TCAACATGCTTTGACT 226
          ||||||||||||||||
Db   171 TTGTGCTGGTATGATT 186

RESULT 6
PCT-US02-38526-589
Sequence 589, Application PC/TUS0238526
GENERAL INFORMATION:
; APPLICANT: Ghosh, Malabika
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Zhao, Qing
; APPLICANT: Xu, Chongjun
; APPLICANT: Mulero, Julio J
; APPLICANT: Boyle, Bryan J.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL POLYPEPTIDES AND POLYNUCL
; FILE REFERENCE: HYS-B1/PCT
; CURRENT APPLICATION NUMBER: PCT/US02/38526
; CURRENT FILING DATE: 2002-12-03
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: PCT/US00/34263
; PRIOR FILING DATE: 2000-12-26
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: PCT/US00/34263
; PRIOR FILING DATE: 2000-12-26
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800

; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: PCT/US01/04098
; PRIOR FILING DATE: 2001-02-05
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 653
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 589
; LENGTH: 1629
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-38526-589

Query Match   8.9%; Score 33.6; DB 1; Length 1629;
Best Local Similarity   52.9%; Pred. No. 1.1;
Matches   72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy   91 TTGAATGTGTTCTGGGAGCTGAAGGATGCACGGTTGTTGTTAAGCCCCTGTTCTTTTCCGTT 150
          ||| ||| |||||||| |||| ||| || |||||||||||| ||| ||||| || |||
Db 1074 TGGAACCCTCTACGCTCTTCGGTTTGGTGGAGTCATCAGATACCTGGGTTATTTCCAGGC 1133

Qy   151 GTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTATGC 210
          ||||||| |||| || |||||||| | ||||| |||||||||||||| ||||||||||||
Db 1134 ATATAATGTGCTGATTTAACAATGCAGCGCCTCACTGCCAAAAAGTTCTTCGGTTTTGTGC 1193

Qy   211 TCAACATGCTTTGACT 226
          ||||||||||||||||
Db 1194 TTGTGCTGGTATGATT 1209

RESULT 7
PCT-US02-38526-587
Sequence 587, Application PC/TUS0238526
GENERAL INFORMATION:
; APPLICANT: Ghosh, Malabika
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Zhao, Qing
; APPLICANT: Xu, Chongjun
; APPLICANT: Mulero, Julio J
; APPLICANT: Boyle, Bryan J.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL POLYPEPTIDES AND POL
; FILE REFERENCE: HYS-B1/PCT
; CURRENT APPLICATION NUMBER: PCT/US02/38526
; CURRENT FILING DATE: 2002-12-03
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: PCT/US00/34263
; PRIOR FILING DATE: 2000-12-26
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: PCT/US01/04098
; PRIOR FILING DATE: 2001-02-05
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 653
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 587
; LENGTH: 1640
; TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1629)
OTHER INFORMATION:
PCT-US02-38526-587

Query Match   8.9%; Score 33.6; DB 1; Length 1640;
Best Local Similarity   52.9%; Pred. No. 1.1;
Matches 72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy   91 TTGAATGTGTGTTCTGGAGCTGAAGGATGCACGGTTGTTAAGCCCCTGTTCTTTTCCGTT 150
         |||||||   |||| |||||| | |  || |  ||  |       |  |||||  |||
Db 1074 TGGAACCTCTACGCCTCTTGGTTTGGGTTGGACTCATCAGATACCTGGGGTTATTTCAGGC 1133

Qy  151 GTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTATGC 210
         ||||    |  ||||     ||||  ||| |  |||| ||||||     |||||||  |||
Db 1134 ATATAAATGTCTGATTTTAACAAATGCAGGCCTCACTGCCCAAAAAGTTCTTCGGTTTTATGC 1193

Qy  211 TCAACATGCTTTGACT 226
         ||  | ||| ||
Db 1194 TTGTGCTGGTATGATT 1209
```

RESULT 8
US-09-513-999C-33882/c
Sequence 33882, Application US/09513999C
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclert, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
FILE REFERENCE: 59.US2.REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 33882
LENGTH: 158
TYPE: DNA
ORGANISM: Homo sapiens
US-09-513-999C-33882

Query Match   8.8%; Score 33.4; DB 5; Length 158;
Best Local Similarity   53.4%; Pred. No. 0.6;
Matches 70; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```
Qy    1 TACGGCTGCTTAAGACGACTTAAGGGGGAATGACGCACGGCCGGCTTAGAGGAACATATCG 60
         | ||||  |||||   || ||  |  ||| |  |  |  |   | |||||| | ||| | |
Db  144 TATAGCTCCTGAGGAGGCCCACAAAGGGAAATCTCCATTCATCTCTTCCTGCCTACAGAAGA 85

Qy   61 AAAACACCCAAGCCGCGGAGTCTCTCACACAAGCTTGAAATCGTTGTTCTGGAGCTGAAGGATGC 120
         |||    ||  | ||| |||| ||| | |  || |||| ||| ||   ||||| |||||| |
Db   84 ACAGCATTAATCCTAGACCCACTAACAGCTCTGGAAGAGTGAATAGGTGGCTGACTGAAGA 25

Qy  121 ACGGTTGTTAA 131
         |  |||| ||
Db   24 AGGGAGAAAA 14
```

RESULT 9
PCT-US02-32700-7
Sequence 7, Application PC/TUS0232700
GENERAL INFORMATION:
APPLICANT: Keith, Tim
APPLICANT: Little, Randall D.
APPLICANT: Van Eerdewegh, Paul
APPLICANT: Dupuis, Josee
APPLICANT: Del Mastro, Richard G.
APPLICANT: Allen, Kristina
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES

TITLE OF INVENTION: RELATING TO RESPIRATORY DISEASES AND OBESITY
FILE REFERENCE: 2976-4045
CURRENT APPLICATION NUMBER: PCT/US02/32700
CURRENT FILING DATE: 2002-10-11
PRIOR APPLICATION NUMBER: 60/328,424
PRIOR FILING DATE: 2001-10-11
OTHER INFORMATION: n = A,T,C or G
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 7
LENGTH: 45450
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(45450)
OTHER INFORMATION: n = A,T,C or G
PCT-US02-32700-7

Query Match   8.6%; Score 32.6; DB 1; Length 45450;
Best Local Similarity   46.6%; Pred. No. 6.7;
Matches 104; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

```
Qy  133 CCCCTGTTCTTTTCCGTTGTTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTAA 192
         | |  ||||| | | | || |||| |  || ||| | ||| || |||||||  ||||| |
Db 2864 CTCATATTTCCTGTTGATGTTTTGTTTTATTGTCCCCCTCACTAAAATTCAGGCTCCTAA 2923

Qy  193 TAGTACTTGGTTTTATTAAGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACAT 252
         |||| |   ||  | || ||   || ||| ||| |  |||| || || | |||||| |||||
Db 2924 CAGTAGGATGCTAATTTCTTGGTATGGTTGAATGAATGAATGAAATTCAATTGGAATGAAC 2983

Qy  253 TGAAGGGATGGTGATGGAGATGCCAATCCATGGAGAATCAGGTGGCGCAGCTATGTTGGTAG 312
         |||||   | || |  |||  || ||  ||||| || || |||   |   |||  | |||
Db 2984 TGAATAAATGAATAAGGAACAGGTGAACTAGATAGGGTTGTTCTGAAGTGATGGGACTAT 3043

Qy  313 CTATAGGCAGAAGTCTTCTTGGCCAAAGAATTCCTCCCGGGGAAGGA 355
         | |  ||  || |  | || ||  || || | ||  | || ||||| |
Db 3044 CCAGAAATGGCTACTCGAGTCACACAAGGTTCTCACTAGATGGA 3086
```

RESULT 10
US-10-271-416-7
Sequence 7, Application US/10271416
GENERAL INFORMATION:
APPLICANT: Keith, Tim
APPLICANT: Little, Randall D.
APPLICANT: Van Eerdewegh, Paul
APPLICANT: Dupuis, Josee
APPLICANT: Del Mastro, Richard G.
APPLICANT: Allen, Kristina
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES
TITLE OF INVENTION: RELATING TO RESPIRATORY DISEASES AND OBESITY
FILE REFERENCE: 2976-4045
CURRENT APPLICATION NUMBER: US/10/271,416
CURRENT FILING DATE: 2002-10-11
PRIOR APPLICATION NUMBER: 60/328,424
PRIOR FILING DATE: 2001-10-11
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 7
LENGTH: 45450
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(45450)
OTHER INFORMATION: n = A,T,C or G
US-10-271-416-7

Query Match   8.6%; Score 32.6; DB 6; Length 45450;
Best Local Similarity   46.6%; Pred. No. 6.7;
Matches 104; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

```
Qy  133 CCCCTGTTCTTTTCCGTTGTTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAG 192
```

```
Db   2864  CTCATATATTCCTGTTGATGTTTTGTTTATTGTCCCCTCACTAAAATTCAGGCTCCCTAA  2923
                 | |  |||  |  ||    |      ||||    |       ||||| | |||
Qy    193  TAGTACTTGGTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACAT   252

Db   2924  CAGTAGGATGCTTATTTCTTGGTATGGTTGAATGAATGAATTCATTTGGAATGAAC      2983
                 |||  |    | ||| || | ||   | |  |  |   |  | ||||||||
Qy    253  TGAAGGATGGTGATGGAGATGCCAATCCATGGAATCAGGTGGGCGCCAGCTTATGTTGGGTAG  312

Db   2984  TGAATAAATGAATAAGGAACAGGTGAACTAGATAGGTTGTTCTTGAAGTGAGTGGACTAT   3043
                 ||   |   |   |||      | ||  ||  ||     ||| || |
Qy    313  CTATAGCACAAGTCTTCTTTGGCAAAGATTCCTCCCGGGGAAGGA                  355

Db   3044  CCAGAAATGGCTACTGAGTGCACACAAGGTTCTCACTAGATGGA                  3086
                 |  |    | | |   ||    ||   ||| || | |||
```

RESULT 11
US-09-531-113-5935
; Sequence 5935, Application US/09531113
; GENERAL INFORMATION:
~ APPLICANT: Byrum, Joseph R.
~ APPLICANT: Heck, Gregory R.
~ APPLICANT: La Rosa, Thomas J.
~ TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
~ TITLE OF INVENTION: Plants
~ FILE REFERENCE: 38-21(15761)B
~ CURRENT APPLICATION NUMBER: US/09/531,113
~ CURRENT FILING DATE: 2000-03-22
~ NUMBER OF SEQ ID NOS: 48629
; SEQ ID NO 5935
; LENGTH: 334
; TYPE: DNA
; ORGANISM: Glycine max
~ OTHER INFORMATION: Clone ID: JC-gmfi022200076b11a1
US-09-531-113-5935

Query Match   8.6%; Score 32.4; DB 5; Length 334;
Best Local Similarity   48.4%;
Matches   90; Conservative   0; Mismatches   96; Indels   0; Gaps   0;

```
Qy   142  TTTTCCGTTGTTGTTTAAATCTAAATGTTCTTTGGAATAAAAACCTCCCGCCAAGTAGTACTTG  201
               ||||  | |||    |   |     || || ||||     ||| |||   |||  ||
Db    18  TATTTGGGTTGTTAGACAACTGGGGTCAGGAAAAGGTGGCCCACTGCCACATTTTGTTAAAG     77

Qy   202  GTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGAGACCTTTGGCACACATTGAAGGGAT    261
               |||  |    | ||| | | ||   | |  |  | |   | | ||||||||
Db    78  GAGAACTCCCTTTAATGTTTTGACTATTAGGAACCTTCCCTGTGAACTCTTTCATGGAGG      137

Qy   262  GGGTGATGGAGATGCCAATCCATGGAATCAGGTGGCGCGCACCTATGTTGGTAGCTATAGCAG   321
               |  |    |  |    | | ||   || |   ||| || | ||| || | |
Db   138  TATGGTGAACATCATATGTCATCATCACTCGGTTTGATCGTATTATTGATAACCATTCTTAA    197

Qy   322  AAGTCT  327
               ||| ||
Db   198  AAGACT  203
```

RESULT 12
US-09-950-084-952/c
; Sequence 952, Application US/09950084
; GENERAL INFORMATION:
~ APPLICANT: George H. Shimer,Jr.
~ APPLICANT: George H. Miller
~ APPLICANT: Roberta S. Hare
~ APPLICANT: Karen J. Shaw
~ TITLE OF INVENTION: Staphylococcus aureus Related Compositions and Methods
~ FILE REFERENCE: 1034/1C963US2
~ CURRENT APPLICATION NUMBER: US/09/950,084
~ CURRENT FILING DATE: 2001-09-10
~ PRIOR APPLICATION NUMBER: US 09/417,811
~ PRIOR FILING DATE: 1999-10-14
~ PRIOR APPLICATION NUMBER: US 09/353,718
~ PRIOR FILING DATE: 1999-07-14
~ PRIOR APPLICATION NUMBER: US 09/266,557
~ PRIOR FILING DATE: 1999-03-11
~ PRIOR APPLICATION NUMBER: US 09/266,556
~ PRIOR FILING DATE: 1999-03-11
~ PRIOR APPLICATION NUMBER: US 09/266,555
~ PRIOR FILING DATE: 1999-03-11
~ PRIOR APPLICATION NUMBER: US 09/266,542
~ PRIOR FILING DATE: 1999-03-11
~ PRIOR APPLICATION NUMBER: US 09/266,541
~ PRIOR FILING DATE: 1999-03-11
~ PRIOR APPLICATION NUMBER: US 09/037,934
~ PRIOR FILING DATE: 1998-03-10
~ PRIOR APPLICATION NUMBER: US 09/036,720
~ PRIOR FILING DATE: 1998-03-06
~ PRIOR APPLICATION NUMBER: US 09/036,338
~ PRIOR FILING DATE: 1998-03-06
~ Remaining Prior Application data removed - See File Wrapper or PALM.
~ NUMBER OF SEQ ID NOS: 7451
; SEQ ID NO 952
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-950-084-952

Query Match   8.5%; Score 32.2; DB 5; Length 264;
Best Local Similarity   61.2%;
Matches   52; Conservative   0; Mismatches   33; Indels   0; Gaps   0;

```
Qy   121  ACGGTTGTTAAGCCCCTGTTCTTTTCCGTTGTTGTTTAAATCTAAATGTTTCTTTGGAATAAAAAC  180
               || ||||||||   | ||| || ||||| | |  |  |   || |  | |||||||| || ||
Db   210  AAGTTTTTTAAGTTTCTCCTTGGTCTCCGTTTGCAAAAATCAAATCATCTTTTGAGCATATAC    151

Qy   181  CTCCCTGCCAAGTAGTACTTGGTTT  205
               | | ||  | | | ||| ||||||
Db   150  ATCTCTACAATGTAATTATTTCTTTT  126
```

RESULT 13
US-09-513-999C-30304/c
; Sequence 30304, Application US/09513999C
; GENERAL INFORMATION:
~ APPLICANT: Dumas Milne Edwards, J.B.
~ APPLICANT: Duclert, A.
~ APPLICANT: Giordano, J.Y.
~ TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
~ FILE REFERENCE: 59.US2.REG
~ CURRENT APPLICATION NUMBER: US/09/513,999C
~ CURRENT FILING DATE: 2000-02-24
~ PRIOR APPLICATION NUMBER: US 60/122,487
~ PRIOR FILING DATE: 1999-02-26
~ NUMBER OF SEQ ID NOS: 36681
~ SOFTWARE: Patent.pm
; SEQ ID NO 30304
; LENGTH: 410
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
~ NAME/KEY: misc_feature
~ LOCATION: 387
~ OTHER INFORMATION: v=a or c or g
; FEATURE:
~ NAME/KEY: misc_feature
~ LOCATION: 388
~ OTHER INFORMATION: n=a, g, c or t
US-09-513-999C-30304

Query Match   8.5%; Score 32; DB 5; Length 410;
Best Local Similarity   51.9%;
Matches   69; Conservative   1; Mismatches   63; Indels   0; Gaps   0;

```
Qy   196  TACTTGGTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGAGACCTTTGGCACACATTGA  255
               ||  |:|   || |||| |  |  | |  ||||    || || || |  | |  | ||
Db   390  TANBTACTCTCATTCTCAAAAGCCAGGCATTCATNAAATGTGGAGCTATGTCTTTTTCTGA  331
```

Qy   256 AGGGATGGTGATGGAGAATGCCAATCATCATGGAATCAGGTGGGCCAGCTATGTGTGGTAGCTA 315
         ||||| || | | || |||| ||| ||||||||| ||| ||||| |||||||||||||||||
Db   330 TGGTAAAGTTCAGGATTTACTAATTGGTACAATCAGAGAGCACTGGTTTCGAGGCTGCTG 271

Qy   316 TAGCAGAAGTCTT 328
         ||||||| |||
Db   270 CAGCAGTACTT 258

RESULT 14
US-10-294-443-103
; Sequence 103, Application US/10294443
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary J.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C12
; CURRENT APPLICATION NUMBER: US/10/294,443
; CURRENT FILING DATE: 2002-11-13
; NUMBER OF SEQ ID NOS: 243
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 103
; LENGTH: 2966
; TYPE: DNA
; ORGANISM: Babesia microti
US-10-294-443-103

Query Match   8.4%;   Score 31.6;   DB 6;   Length 2966;
Best Local Similarity 51.4%; Conservative 0; Mismatches 69; Indels 0;
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy   142 TTTTCCGTTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTG 201
         | |||||||||||||||||| |||||||| ||||||||||||||||||||||||||||||
Db   271 TATTCTACTTTTTAATATAATTTTTTTATTCAATAATAATACTCTTTCATAATTTGTACTAT 330

Qy   202 GTTTTATGCTCAACATGCTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGAT 261
         || ||||||||||||||||||| ||| |||||||||||||| ||| ||||| ||
Db   331 TTTATATAATCATATATATATATAATTAATTGATATATATCAATATTTTGTACTAT 390

Qy   262 GGTGATGGAGATGCCAATCCAT 283
         ||||| |||| || || ||||
Db   391 GATGATATACATGAAATATGCAT 412

RESULT 15
US-10-294-443-112
; Sequence 112, Application US/10294443
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary J.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C12
; CURRENT APPLICATION NUMBER: US/10/294,443
; CURRENT FILING DATE: 2002-11-13
; NUMBER OF SEQ ID NOS: 243
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 112
; LENGTH: 3011
; TYPE: DNA
; ORGANISM: Babesia microti

US-10-294-443-112

Query Match   8.4%;   Score 31.6;   DB 6;   Length 3011;
Best Local Similarity 51.4%; Conservative 0; Mismatches 69; Indels 0;
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy   142 TTTTCCGTTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTG 201
         | |||||||||||||||||| |||||||| ||||||||||||||||||||||||||||||
Db   271 TATTCTACTTTTTAATATAATTTTTTTATTCAATAATAATACTCTTTCATAATTTGTACTAT 330

Qy   202 GTTTTATGCTCAACATGCTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGAT 261
         || ||||||||||||||||||| ||| |||||||||||||| ||| ||||| ||
Db   331 TTTATATAATCATATATATATATAATTAATTGATATATATCAATATTTTGTACTAT 390

Qy   262 GGTGATGGAGATGCCAATCCAT 283
         ||||| |||| || || ||||
Db   391 GATGATATACATGAAATATGCAT 412

Search completed: January 31, 2003, 01:19:25
Job time : 239 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: January 31, 2003, 01:12:19 ; Search time 3488 Seconds
(without alignments)
2724.731 Million cell updates/sec

Title: US-09-803-719-222
Perfect score: 378
Sequence: 378
1 tacggctgcttaagacgact..........taccattggagaaccatgcg 378

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database : Pending_Patents_NA_Main:*

```
 1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq:*
 2: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
 3: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
 4: /cgn2_6/ptodata/1/pna/US080_COMB.seq:*
 5: /cgn2_6/ptodata/1/pna/US081_COMB.seq:*
 6: /cgn2_6/ptodata/1/pna/US082_COMB.seq:*
 7: /cgn2_6/ptodata/1/pna/US083_COMB.seq:*
 8: /cgn2_6/ptodata/1/pna/US084_COMB.seq:*
 9: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US093_COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US094_COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US096A_COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US096B_COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US096C_COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US096D_COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US096E_COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US097A_COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US097B_COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US097C_COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US098A_COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US098B_COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US098C_COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US099A_COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US099B_COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US099C_COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US099D_COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US100A_COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US100B_COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US100C_COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US101A_COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US101B_COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq:*
44: /cgn2_6/ptodata/1/pna/US6000_COMB.seq:*
45: /cgn2_6/ptodata/1/pna/US6001_COMB.seq:*
46: /cgn2_6/ptodata/1/pna/US6002_COMB.seq:*
47: /cgn2_6/ptodata/1/pna/US6003_COMB.seq:*
48: /cgn2_6/ptodata/1/pna/US6004_COMB.seq:*
49: /cgn2_6/ptodata/1/pna/US6005_COMB.seq:*
50: /cgn2_6/ptodata/1/pna/US6006_COMB.seq:*
51: /cgn2_6/ptodata/1/pna/US6007_COMB.seq:*
52: /cgn2_6/ptodata/1/pna/US6008_COMB.seq:*
53: /cgn2_6/ptodata/1/pna/US6009_COMB.seq:*
54: /cgn2_6/ptodata/1/pna/US6010_COMB.seq:*
55: /cgn2_6/ptodata/1/pna/US6011_COMB.seq:*
56: /cgn2_6/ptodata/1/pna/US6012_COMB.seq:*
57: /cgn2_6/ptodata/1/pna/US6013_COMB.seq:*
58: /cgn2_6/ptodata/1/pna/US6014_COMB.seq:*
59: /cgn2_6/ptodata/1/pna/US6015_COMB.seq:*
60: /cgn2_6/ptodata/1/pna/US6016_COMB.seq:*
61: /cgn2_6/ptodata/1/pna/US6017_COMB.seq:*
62: /cgn2_6/ptodata/1/pna/US6018_COMB.seq:*
63: /cgn2_6/ptodata/1/pna/US6019_COMB.seq:*
64: /cgn2_6/ptodata/1/pna/US6020_COMB.seq:*
65: /cgn2_6/ptodata/1/pna/US6021_COMB.seq:*
66: /cgn2_6/ptodata/1/pna/US6022_COMB.seq:*
67: /cgn2_6/ptodata/1/pna/US6023_COMB.seq:*
68: /cgn2_6/ptodata/1/pna/US6024_COMB.seq:*
69: /cgn2_6/ptodata/1/pna/US6025_COMB.seq:*
70: /cgn2_6/ptodata/1/pna/US6026_COMB.seq:*
71: /cgn2_6/ptodata/1/pna/US6027_COMB.seq:*
72: /cgn2_6/ptodata/1/pna/US6028_COMB.seq:*
73: /cgn2_6/ptodata/1/pna/US6029_COMB.seq:*
74: /cgn2_6/ptodata/1/pna/US6030_COMB.seq:*
75: /cgn2_6/ptodata/1/pna/US6031_COMB.seq:*
76: /cgn2_6/ptodata/1/pna/US6032_COMB.seq:*
77: /cgn2_6/ptodata/1/pna/US6033_COMB.seq:*
78: /cgn2_6/ptodata/1/pna/US6034_COMB.seq:*
79: /cgn2_6/ptodata/1/pna/US6035_COMB.seq:*
80: /cgn2_6/ptodata/1/pna/US6036_COMB.seq:*
81: /cgn2_6/ptodata/1/pna/US6037_COMB.seq:*
82: /cgn2_6/ptodata/1/pna/US6038_COMB.seq:*
83: /cgn2_6/ptodata/1/pna/US6039_COMB.seq:*
84: /cgn2_6/ptodata/1/pna/US6040_COMB.seq:*
85: /cgn2_6/ptodata/1/pna/US6041_COMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB ID	Description
	1	378	100.0	378	31	US-09-803-719-222 Sequence 222, App
	2	342	90.5	512	24	US-09-621-976-3594 Sequence 3594, App
	3	342	90.5	512	32	US-09-834-366-8066 Sequence 8066, Ap
	4	342	90.5	512	58	US-60-147-873-3594 Sequence 3594, Ap
	5	342	90.5	512	63	US-60-197-873-8066 Sequence 8066, Ap
	6	299.6	79.3	327	31	US-09-803-719-276 Sequence 276, App
	7	279.6	74.0	322	31	US-09-803-719-295 Sequence 295, App
	8	275	72.8	318	31	US-09-718-457-565 Sequence 565, App
	9	226.8	60.0	562	28	US-09-718-457-565 Sequence 565, App
c	10	131	34.7	403	20	US-09-534-844A-5172 Sequence 5172, Ap
c	11	83.4	22.1	524	27	US-09-689-907-861 Sequence 861, Ap
c	12	79.4	21.0	806	39	US-10-098-754-19036 Sequence 19036, A
	13	75.6	20.0	125	27	US-09-685-045-1053 Sequence 1053, Ap
	14	48.8	12.9	381	17	US-09-362-510A-51216 Sequence 51216, A
	15	48.8	12.9	381	17	US-09-362-510A-51216 Sequence 51216, A
	16	40.2	10.6	453	34	US-09-904-013-51216 Sequence 51216, A
	17	40.2	10.6	453	17	US-09-306-350A-28514 Sequence 28514, A
	18	39.8	10.5	613	25	US-09-909-629-28514 Sequence 28514, A
c	19	39.2	10.4	1266	19	US-09-652-123-5950 Sequence 5950, A
c	20	39.2	10.4	1266	25	US-09-652-123-14754 Sequence 14754, A
c	21	39.2	10.4	1266	31	US-09-819-091A-14754 Sequence 14754, A

22	39.2	10.4	1408	35	US-09-938-842A-3096	Sequence 3096, Ap
23	39.2	10.4	91071	20	US-09-554-859-363	Sequence 363, App
24	39.2	10.4	91071	31	US-09-803-736-363	Sequence 363, App
25	36	9.5	361	22	US-09-552-086-14635	Sequence 14635,
26	36	9.3		33	US-09-874-770A-83442	Sequence 83442, A
27	35.2	9.3	731	65	US-60-211-750-82309	Sequence 82309, A
28	35.2	9.3	1671	1	PCT-US01-10366-23	Sequence 23, Appl
29	35.2	9.3	1671	40	US-10-114-153-23	Sequence 23, Appl
30	35.2	9.3	1677	1	PCT-US01-10366-21	Sequence 21, Appl
31	35.2	9.3	1677	40	US-10-114-153-21	Sequence 21, Appl
32	35.2	9.3	2067	1	PCT-US01-10366-27	Sequence 27, Appl
33	35.2	9.3	2067	40	US-10-114-153-27	Sequence 27, Appl
34	35.2	9.3	2130	1	PCT-US01-10366-25	Sequence 25, Appl
35	35.2	9.3	2130	40	US-10-114-153-25	Sequence 25, Appl
36	35	9.3	315	39	US-10-098-754-7989	Sequence 7989, Ap
37	35	9.3	1596	24	US-09-634-306B-255548	Sequence 255548,
38	35	9.3	1596	38	US-10-027-632-256548	Sequence 255548,
39	34.6	9.3	1596	24	US-09-634-306B-256549	Sequence 256549,
40	34.6	9.3	1596	38	US-10-027-632-256549	Sequence 256549,
41	34.6	9.2	64407	24	US-09-620-392-30211	Sequence 30211, A
42	34.6	9.2	102394	28	US-09-702-134-4852	Sequence 4852, Ap
43	34.6	9.2	102394	31	US-09-815-264-82461	Sequence 82461, A
44	34.2	9.0	588	33	US-09-865-439A-96308	Sequence 96308, A
45	34.2	9.0	588	64	US-60-207-458-140229	Sequence 140229,

RESULT 1
US-09-719-222
Sequence 222, Application US/09803719
GENERAL INFORMATION:
APPLICANT: Williams, Lewis T.
APPLICANT: Escobedo, Jaime
APPLICANT: Innis, Michael A.
APPLICANT: Garcia, Pablo Dominguez
APPLICANT: Sudduth-Klinger, Julie
APPLICANT: Reinhard, Christoph
APPLICANT: Giese, Klaus
APPLICANT: Randazzo, Filippo
APPLICANT: Kennedy, Giulia C.
APPLICANT: Pot, David
APPLICANT: Leshkowitz, Dena
APPLICANT: Kita, David
APPLICANT: Kassam, Altaf
APPLICANT: Lamson, George
APPLICANT: Drmanac, Radoje
APPLICANT: Crkvenjakov, Radomir
APPLICANT: Dickson, Mark
APPLICANT: Drmanac, Snezana
APPLICANT: Labat, Ivan
APPLICANT: Jones, Lee William
APPLICANT: Garcia, Veronica
APPLICANT: Stache-Crain, Birgit
TITLE OF INVENTION: Human Genes and Gene Products
FILE REFERENCE: 1624.002
CURRENT APPLICATION NUMBER: US-09/803,719
CURRENT FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/188,609
PRIOR FILING DATE: 2000-03-09
NUMBER OF SEQ ID NOS: 2396
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 222
LENGTH: 378
TYPE: DNA
ORGANISM: Homo sapiens
US-09-803-719-222

Query Match 100.0%; Score 378; DB 31; Length 378;
Best Local Similarity 100.0%; Pred. No. 3.8e-106;
Matches 378; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ALIGNMENTS

```
Qy     1 TACGGCTGCTTAAGACGACTTAAGGGGGGAATGACGCAGCGAGCGCGGCTCTTAGAGGAACATATGG  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 TACGGCTGCTTAAGACGACTTAAGGGGGGAATGACGCAGCGAGCGCGGCTCTTAGAGGAACATATGG  60

Qy    61 AAAACACCCAAGCCGGAGTCTCTCACAAAGCTTGAATGTGTGTTGTTCTGGAGCTGAAGGATGC 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AAAACACCCAAGCCGGAGTCTCTCACAAAGCTTGAATGTGTGTTGTTCTGGAGCTGAAGGATGC 120

Qy   121 ACGGTTGTTAAGCCCCCTGTTCTTTCGTTGTTTTAATCTAAGTGTTGGAATAAAAAC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ACGGTTGTTAAGCCCCCTGTTCTTTCGTTGTTTTAATCTAAGTGTTGGAATAAAAAC 180

Qy   181 CTCCCCTGCCCAAGTAGTACTTGGTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGACC 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTCCCCTGCCCAAGTAGTACTTGGTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGACC 240

Qy   241 TTTGGCACACATTGAAGGGATGGTGATGGGAGAGATGCCAATCACCATGGAATCAGGTGGGCGCAG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TTTGGCACACATTGAAGGGATGGTGATGGGAGAGATGCCAATCACCATGGAATCAGGTGGGCGCAG 300

Qy   301 CTATGTTGGTAGCTATAGCAGAGAAGTCTTCTTGGCAAAGATTCCTCCCGGGAACGGAAGGTA 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CTATGTTGGTAGCTATAGCAGAGAAGTCTTCTTGGCAAAGATTCCTCCCGGGAACGGAAGGTA 360

Qy   361 CCATTGGAGAACCATGCG 378
           ||||||||||||||||||
Db   361 CCATTGGAGAACCATGCG 378
```

RESULT 2
US-09-621-976-3594
Sequence 3594, Application US/09621976
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 3594
LENGTH: 512
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 84..404
US-09-621-976-3594

Query Match 90.5%; Score 342; DB 24; Length 512;
Best Local Similarity 98.6%; Pred. No. 6.3e-95;
Matches 345; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy    28 GAATGACGCAGCGAGCGCGGCTCTTAGAGGAACATATGGAAAAACACCCAAGCCGGAGTCTCTCACA 87
           |||||||||||| | ||||   |||||||||||||  |||||| ||||||||||||||||||||||
Db    17 GAATGACGCAGCAGCTCTGAGAGGAACATAAGGAAAAACACCCAAGCCGGAGTCTCTCACA 76

Qy    88 AGCTTGAATGTGTGTTCTGGAGCTGAACGATGCACGGTTGTTGAAGCCCCTGTTCTTTCC 147
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    77 AGCTTGAATGTGTGTTCTGGAGCTGAACGATGCACGGTTGTTGAAGCCCCTGTTGTTTTCC 136

Qy   148 GTTGTTTTAATCTAATGTTCTTTTGGAATAAAAACCTCCCTGCCCAAGTAGTACTTGGTTTTA 207
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   137 GTTGTTTTAATCTAATGTTCTTTTGGAATAAAAACCTCCCTGCCCAAGTAGTACTTGGTTTTA 196

Qy   208 TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 267
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   197 TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 256

Qy   268 GGAGATGCCAATCCATGGAATCAGGTGCGCAGCTATGTTGGTAGCTATAGCAGAGAAGTCT 327
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   268 GGAGATGCCAATCCATGGAATCAGGTGCGCAGCTATGTTGGTAGCTATAGCAGAGAAGTCT 327
```

Db   257 GGAGATGCCAATCCATGGAATCAGGTGGCACAGCTANGTTGGTTAGCTATAGCAGAAGTCT 316

Qy   328 TCTTGGCAAAGAGATTCCTCCCGGGAAGGAAGGTACCATTGGAGAACCATGC 377
          ||||||||||||| ||||||||| ||||||||||||||||||||||||||||
Db   317 TCTTGGCAAAAGATTCCTCCAGGGAAGGAAGGTACCATTGGAGAACCATGC 366

RESULT 3
US-09-834-366-8066
Sequence 8066, Application US/09834366
; GENERAL INFORMATION:
;   APPLICANT: Bejanin, Stephane
;   APPLICANT: Tanaka, Hiroaki
;   APPLICANT: Dumas Milne Edwards, Jean Baptiste
;   APPLICANT: Jobert, Severin
;   APPLICANT: Giordano, Jean-Yves
;   TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;   FILE REFERENCE: 81.US2.REG
;   CURRENT APPLICATION NUMBER: US/09/834,366
;   CURRENT FILING DATE: 2001-04-13
;   PRIOR APPLICATION NUMBER: US 60/197,873
;   PRIOR FILING DATE: 2000-04-18
;   NUMBER OF SEQ ID NOS: 52153
;   SOFTWARE: Patent.pm
; SEQ ID NO 8066
;   LENGTH: 512
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 84..404
US-09-834-366-8066

Query Match   90.5%; Score 342; DB 32; Length 512;
Best Local Similarity   98.6%; Pred. No. 6.3e-95;
Matches 345; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

Qy   28 GAATGACGCAGCGCGGCTCTTAGAGGAACATATGGAAAAACACCCAAGCCCGGAGTCTCACA 87
         |||||||||||||| ||||| |||||||||||||||||||||||||||||||||||||||
Db   17 GAATGACGCAGCAGCTCTGAGAGGAACATAAGGAAAAACACCCAAGCCGGAGTCTCTCACA 76

Qy   88 AGCTTGAAATGTGTTCTGGAGCTGAAAGAATGCACGGTTGTTAAGCCCCTGTTCTTTTTCC 147
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   77 AGCTTGAAATGTGTTCTGGAGCTGAAAGAATGCACGGTTGTTAAGCCCCTGTTCTTTTTCC 136

Qy   148 GTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCCAAGTAGTACTTGGTTTTA 207
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   137 GTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCCAAGTAGTACTTGGTTTTA 196

Qy   208 TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 267
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   197 TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 256

Qy   268 GGAGATGCCAATCCATGGAATCAGGTGGCGCCAGCTATGTTGGTTAGCTATAGCAGAAGTCT 327
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   257 GGAGATGCCAATCCATGGAATCAGGTGGCACAGCTATGTTGGTTAGCTATAGCAGAAGTCT 316

Qy   328 TCTTGGCAAAGAGATTCCTCCCGGGAAGGAAGGTACCATTGGAGAACCATGC 377
         ||||||||||||| ||||||||| ||||||||||||||||||||||||||||
Db   317 TCTTGGCAAAAGATTCCTCCAGGGAAGGAAGGTACCATTGGAGAACCATGC 366

RESULT 4
US-60-147-499-3594
Sequence 3594, Application US/6014749
; GENERAL INFORMATION:
;   APPLICANT: Dumas Milne Edwards, J.B.
;   APPLICANT: Jobert, S.
;   APPLICANT: Giordano, J.Y.
;   TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;   FILE REFERENCE: GENSET.054PR2
;   CURRENT APPLICATION NUMBER: US/60/147,499
;   CURRENT FILING DATE: 1999-08-05

;   NUMBER OF SEQ ID NOS: 19335
;   SOFTWARE: Patent.pm
; SEQ ID NO 3594
;   LENGTH: 512
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 84..404
US-60-147-499-3594

Query Match   90.5%; Score 342; DB 58; Length 512;
Best Local Similarity   98.6%; Pred. No. 6.3e-95;
Matches 345; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

Qy   28 GAATGACGCAGCGCGGCTCTTAGAGGAACATATGGAAAAACACCCAAGCCCGGAGTCTCACA 87
         |||||||||||||| ||||| |||||||||||||||||||||||||||||||||||||||
Db   17 GAATGACGCAGCAGCTCTGAGAGGAACATAAGGAAAAACACCCAAGCCGGAGTCTCTCACA 76

Qy   88 AGCTTGAAATGTGTTCTGGAGCTGAAAGAATGCACGGTTGTTAAGCCCCTGTTCTTTTTCC 147
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   77 AGCTTGAAATGTGTTCTGGAGCTGAAAGAATGCACGGTTGTTAAGCCCCTGTTCTTTTTCC 136

Qy   148 GTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCCAAGTAGTACTTGGTTTTA 207
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   137 GTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCCAAGTAGTACTTGGTTTTA 196

Qy   208 TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 267
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   197 TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 256

Qy   268 GGAGATGCCAATCCATGGAATCAGGTGGCGCCAGCTATGTTGGTTAGCTATAGCAGAAGTCT 327
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   257 GGAGATGCCAATCCATGGAATCAGGTGGCACAGCTATGTTGGTTAGCTATAGCAGAAGTCT 316

Qy   328 TCTTGGCAAAGAGATTCCTCCCGGGAAGGAAGGTACCATTGGAGAACCATGC 377
         ||||||||||||| ||||||||| ||||||||||||||||||||||||||||
Db   317 TCTTGGCAAAAGATTCCTCCAGGGAAGGAAGGTACCATTGGAGAACCATGC 366

RESULT 5
US-60-197-873-8066
Sequence 8066, Application US/60197873
; GENERAL INFORMATION:
;   APPLICANT: Bejanin, Stephane
;   APPLICANT: Tanaka, Hiroaki
;   APPLICANT: Dumas Milne Edwards, Jean Baptiste
;   APPLICANT: Jobert, Severin
;   APPLICANT: Giordano, Jean-Yves
;   TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;   FILE REFERENCE: 81.US1.PRO
;   CURRENT APPLICATION NUMBER: US/60/197,873
;   CURRENT FILING DATE: 2000-04-18
;   NUMBER OF SEQ ID NOS: 52153
;   SOFTWARE: Patent.pm
; SEQ ID NO 8066
;   LENGTH: 512
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 84..404
US-60-197-873-8066

Query Match   90.5%; Score 342; DB 63; Length 512;
Best Local Similarity   98.6%; Pred. No. 6.3e-95;
Matches 345; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

Qy   28 GAATGACGCAGCGCGGCTCTTAGAGGAACATATGGAAAAACACCCAAGCCCGGAGTCTCACA 87
         |||||||||||||| ||||| |||||||||||||||||||||||||||||||||||||||
Db   17 GAATGACGCAGCAGCTCTGAGAGGAACATAAGGAAAAACACCCAAGCCGGAGTCTCTCACA 76

Qy   88 AGCTTGAAATGTGTTCTGGAGCTGAAAGAATGCACGGTTGTTAAGCCCCTGTTCTTTTTCC 147

Left column — RESULT 6

```
RESULT 6
US-09-803-719-276
Sequence 276, Application US/09803719
GENERAL INFORMATION:
  APPLICANT: Williams, Lewis T.
  APPLICANT: Escobedo, Jaime
  APPLICANT: Innis, Michael A.
  APPLICANT: Garcia, Pablo Dominguez
  APPLICANT: Sudduth-Klinger, Julie
  APPLICANT: Reinhard, Christoph
  APPLICANT: Giese, Klaus
  APPLICANT: Randazzo, Filippo
  APPLICANT: Kennedy, Giulia C.
  APPLICANT: Pot, David
  APPLICANT: Kassam, Altaf
  APPLICANT: Lamson, George
  APPLICANT: Drmanac, Radoje
  APPLICANT: Crkvenjakov, Radomir
  APPLICANT: Dickson, Mark
  APPLICANT: Drmanac, Snezana
  APPLICANT: Labat, Ivan
  APPLICANT: Leshkowitz, Dena
  APPLICANT: Kita, David
  APPLICANT: Jones, Veronica
  APPLICANT: Garcia, Lee William
  APPLICANT: Stache-Crain, Birgit
  FILE REFERENCE: 1624.002
  TITLE OF INVENTION: Human Genes and Gene Products
  CURRENT APPLICATION NUMBER: US/09/803,719
  CURRENT FILING DATE: 2001-03-09
  PRIOR APPLICATION NUMBER: 60/188,609
  PRIOR FILING DATE: 2000-03-09
  NUMBER OF SEQ ID NOS: 2396
  SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 276
  LENGTH: 327
  TYPE: DNA
  ORGANISM: Homo sapiens

Query Match       79.3%;  Score 299.6;  DB 31;  Length 327;
Best Local Similarity 98.7%;  Pred. No. 7.5e-82;
Matches 302;  Conservative 0;  Mismatches 4;  Indels 0;  Gaps 0;

US-09-803-719-276

Qy  148  GTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA  207
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81  AGCTTGAAATGTGTGTTCTGGAGCTGAAGGATGCACGGTTGTTAAGCCCCTGTTCTTTTCC  140

Qy   88  AGCTTGAATGTGTGTTCTGGAGCTGAAGGATGCACGGTTGTTAAGCCCCTGTTCTTTTCC  147
         |||||||||||||||||||||||||||||||||||||||||||||||||||| || ||||
Db   21  GAATGACGCAGCAGCTCTGAGAGGAACATAAGGAAAACACCCAAGCCGGAGTCTCTCACA  80

Qy   28  GAATGACGCAGCAGCTCTTAGAGGAACATATGGAAAACACCCAAGCCGGAGTCTCTCACA  87
         |||||||||||||||||||| ||||||||| |||||||||||||||||||||||||||||
Db  137  GTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA  196

Qy  208  TGCTCAACATGCTTTGACTGTTGTGAAAAGAGACCTTTGGCACACATTGAAAGGATGGTGAT  267
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  197  TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAAGGGATGGTGAT  256

Qy  268  GGAGATGCCAATCCATGGAATCAGGTGGCGCCGAGCTATGTGTGTTGGTAGCTATAGCAGAAGTCT  327
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  257  GGAGATGCCAATCCATGGAATCAGGTGGCACAGCTATGTTGGGTAGCTATAGCAGAAGTCT  316

Qy  328  TCTTGGCAAAGATTCCTCCCGGGAAGGAAGGTACCATTGGAGAACCATGC  377
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  317  TCTTGGCAAAGATTCCTCCAGGGAAGGAAGGTACCATTGGAGAAACCATGC  366
```

Right column — RESULT 7

```
RESULT 7
US-09-803-719-295
Sequence 295, Application US/09803719
GENERAL INFORMATION:
  APPLICANT: Williams, Lewis T.
  APPLICANT: Escobedo, Jaime
  APPLICANT: Innis, Michael A.
  APPLICANT: Garcia, Pablo Dominguez
  APPLICANT: Sudduth-Klinger, Julie
  APPLICANT: Reinhard, Christoph
  APPLICANT: Giese, Klaus
  APPLICANT: Randazzo, Filippo
  APPLICANT: Kennedy, Giulia C.
  APPLICANT: Pot, David
  APPLICANT: Kassam, Altaf
  APPLICANT: Lamson, George
  APPLICANT: Drmanac, Radoje
  APPLICANT: Crkvenjakov, Radomir
  APPLICANT: Dickson, Mark
  APPLICANT: Drmanac, Snezana
  APPLICANT: Labat, Ivan
  APPLICANT: Leshkowitz, Dena
  APPLICANT: Kita, David
  APPLICANT: Jones, Veronica
  APPLICANT: Garcia, Lee William
  APPLICANT: Stache-Crain, Birgit
  FILE REFERENCE: 1624.002
  TITLE OF INVENTION: Human Genes and Gene Products
  CURRENT APPLICATION NUMBER: US/09/803,719
  CURRENT FILING DATE: 2001-03-09
  PRIOR APPLICATION NUMBER: 60/188,609
  PRIOR FILING DATE: 2000-03-09
  NUMBER OF SEQ ID NOS: 2396
  SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 295
  LENGTH: 322
  TYPE: DNA
  ORGANISM: Homo sapiens
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (1)...(322)
  OTHER INFORMATION: n = A,T,C or G

Query Match       74.0%;  Score 279.6;  DB 31;  Length 322;
Best Local Similarity 98.6%;  Pred. No. 1.2e-75;
Matches 282;  Conservative 0;  Mismatches 4;  Indels 0;  Gaps 0;

US-09-803-719-295

Qy  148  GTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA  207
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   96  AGCTTGAATGTGTGTTCTGGAGCTGAAGGATGCACGGTTGTTAAGCCCCTGTTCTTTTCC  155

Qy   88  AGCTTGAATGTGTGTTCTGGAGCTGAAGGATGCACGGTTGTTAAGCCCCTGTTCTTTTCC  147
         ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   36  GAATGACGCAGCAGCTCTGAGAGGAACATAAGGAAAACACCCAAGCCGGAGTCTCTCACA  95

Qy   28  GAATGACGCAGCAGCTCTTAGAGGAACATATGGAAAACACCCAAGCCGGAGTCTCTCACA  87
         |||||||||||||||||||| ||||||||| |||||||||||||||||||||||||||||
Db  141  GTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA  200

Qy  208  TGCTCAACATGCTTTGACTGTTGTGAAAAGAGACCTTTGGCACACATTGAAAGGGATGGTGAT  267
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  201  TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAAGGGATGGTGAT  260

Qy  268  GGAGATGCCAATCCATGGAATCAGGTGGCGCCGAGCTATGTGTGTTGGTAGCTATAGCAGAAGTCT  327
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  261  GGAGATGCCAATCCATGGAATCAGGTGGCACAGCTATGTTGGGTAGCTATAGCAGAAGTCT  320

Qy  328  TCTTGG  333
         ||||||
Db  321  TCTTGG  326
```

```
Db    156 GTTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA 215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    208 TGCTCAACATGCTTTGACTGTTGAAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    216 TGCTCAACATGCCTTTGACTGTTGAAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 275
Qy    268 GGAGATGCCAATCCATGGAATCAGGTGGCGCAGCTATGTGGTAGC 313
          |||||||||||||||||||||||||||| |||||||||||||||
Db    276 GGAGATGCCAATCCATGGAATCAGGTGGCACAGCTATGTGGTAGC 321
```

RESULT 8
US-09-803-719-300
Sequence 300, Application US/09803719
GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Escobedo, Jaime
; APPLICANT: Innis, Michael A.
; APPLICANT: Garcia, Pablo Dominguez
; APPLICANT: Reinhard, Christoph
; APPLICANT: Sudduth-Klinger, Julie
; APPLICANT: Giese, Klaus
; APPLICANT: Randazzo, Filippo
; APPLICANT: Kennedy, Giulia C.
; APPLICANT: Pot, David
; APPLICANT: Kassam, Altaf
; APPLICANT: Lamson, George
; APPLICANT: Drmanac, Radoje
; APPLICANT: Crkvenjakov, Radomir
; APPLICANT: Dickson, Mark
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Garcia, Veronica
; APPLICANT: Jones, Lee William
; APPLICANT: Stache-Crain, Birgit
; TITLE OF INVENTION: Human Genes and Gene Products
; FILE REFERENCE: 1624.002
; CURRENT APPLICATION NUMBER: US/09/803,719
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,609
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 2396
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 300
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-803-719-300

Query Match   72.8%; Score 275; DB 31; Length 318;
Best Local Similarity   98.2%; Pred. No. 3.2e-74;
Matches 278; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

```
Qy     28 GAATGACGCACGCGGCTCTTAGAGGACACATATGGAAAAACACCCAAGCCGGAGTCTCACA 87
          ||||||||| | | ||||| |||||||| |||||||||||||||||||| |||||||||||
Db     36 GAATGACGCAGCGCTCGAGAGGAACATAAGGAAAAACACCCAAGCCGAGTCTCACA 95
Qy     88 AGCTTGAATGCGTCGTTCTGGAGCTGAAGGATGCACGCGGTGTTGTTAAGCCCCTGTTTCC 147
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db     96 AGCTTGAATGTGTGTCTCGAGCTGAAGGATGCACGCGGTGTTGTTAAGCCCCTGTTTCTTTCC 155
Qy    148 GTTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA 207
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    156 GTTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA 215
Qy    208 TGCTCAACATGCTTTGACTGTTGAAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 267
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    216 TGCTCAACATGCTTTGACTGTTGAAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 275
Qy    268 GGAGATGCCAATCCATGGAATCAGGTGGCGCGGAGCTATGTTGGGT 310
```

```
Db    276 GGAGATGCCAATCCATGGAATCAGGTGGCACAGCTATGTGGT 318
```

RESULT 9
US-09-718-457-565
Sequence 565, Application US/09718457
GENERAL INFORMATION:
; APPLICANT: Wattler, Sigrid
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Novel Human Polynucleotides and
; Polypeptides Encoded Thereby
; FILE REFERENCE: LEX-0093-USA
; CURRENT APPLICATION NUMBER: US/09/718,457
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: US 60/167,870
; PRIOR FILING DATE: 1999-11-29
; NUMBER OF SEQ ID NOS: 1008
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 565
; LENGTH: 562
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-718-457-565

Query Match   60.0%; Score 226.8; DB 28; Length 562;
Best Local Similarity   99.1%; Pred. No. 3.5e-59;
Matches 228; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

```
Qy    148 GTTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA 207
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA 60
Qy    208 TGCTCAACATGCTTTGACTGTTGAAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGCTCAACATGCTTTGACTGTTGAAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 120
Qy    268 GGAGATGCCAATCCATGGAATCAGGTGGCGCAGCTATGTTGGTAGCTGTAGCTATAGCAGAGAAGTCT 327
          |||||||||||||||||||||||||||| |||||||| ||||||||| ||||||||||||||||
Db    121 GGAGATGCCAATCCATGGAATCAGGTGGCACAGCTATGTTGGTAGCTATAGCAGAGAAGTCT 180
Qy    328 TCTTGGCAAAGATTCCTCCCGGGGAGGAAGGTTACCATTGGAGAACCATGC 377
          |||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    181 TCTTGGCAAAGATTCCTCCCAGGGGAGGAAGGTTACCATTGGAGAACCATGC 230
```

RESULT 10
US-09-534-844A-5172
Sequence 5172, Application US/09534844A
GENERAL INFORMATION:
; APPLICANT: Dramanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark C.
; APPLICANT: Jones, Lee W.
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
; From Various cDNA Libraries
; FILE REFERENCE: 21272-109 (775)
; CURRENT APPLICATION NUMBER: US/09/534,844A
; CURRENT FILING DATE: 2001-03-24
; PRIOR APPLICATION NUMBER: 60/126,605
; PRIOR FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 14510
; SOFTWARE: Hy-patent.pl Version 1.1
; SEQ ID NO 5172
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-534-844A-5172

Query Match   34.7%; Score 131; DB 20; Length 403;
Best Local Similarity   95.7%; Pred. No. 1.6e-29;

```
Matches   156;  Conservative   0;  Mismatches   5;  Indels   2;  Gaps   2;

Qy  148 GTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCCGCCAAGCTAGTAGTACTTGGTTTTA 207
        ||||||| |||||||||||||||||||||||||||||||   ||||||||||  |||||  |||| 
Db  242 GGTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCCCAAGGAGTACTT--GAITTA 300

Qy  208 TGCTCAACANTGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 267
        ||||||||| |||||||||||||||||||||||||||||||||||||||||||||| -||||| 
Db  301 TGCTCAACAATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGAT-GTGAT 359

Qy  268 GGAGATGCCAAATCCATGGAAATCAGGTGGCGCCAGCCTATGTTGGT 310
        |||||||||| |||||||||| ||||  |||||||| |||||||||| 
Db  360 GGAGATGCCAAATCCATGGAAATCAAGTGGCACAGCCTAIGTTGGT 402
```

RESULT 11
US-09-689-907-861/c
Sequence 861, Application US/09689907
; GENERAL INFORMATION:
; APPLICANT: Nehls, Michael
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Novel Human Polynucleotides and the
; TITLE OF INVENTION: Polypeptides Encoded Thereby
; FILE REFERENCE: LEX-0064-USA
; CURRENT APPLICATION NUMBER: US/09/689,907
; CURRENT FILING DATE: 2000-10-11
; PRIOR APPLICATION NUMBER: US 60/158,799
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 1508
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 861
; LENGTH: 524
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-689-907-861

```
Query Match   22.1%;  Score 83.4;  DB 27;  Length 524;
Best Local Similarity   98.8%;  Pred. No. 1e-14;
Matches   84;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy  262 GGTGATGGAGATGCCAATCCATGGAAATCAGGTGGCGCCAGCCTATGTTGGTGGTAGCTATAGCAG 321
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db  249 GGTGATGGAGATGCCAATCCATGGAAATCAGGTGGCGCCAGCCTATGTTGGTGGTAGCTATAGCAG 190

Qy  322 AAGTCTTCTTTGGCAAAGATTCCTCC 346
        |||||||||||||||||||||||||| 
Db  189 AAGTCTTCTTTGGCAAAGATTCCTCC 165
```

RESULT 12
US-10-098-754-19036/c
Sequence 19036, Application US/10098754
; GENERAL INFORMATION:
; APPLICANT: Harrington, John J.
; APPLICANT: Jackson, P. David
; APPLICANT: Sherf, Bruce A.
; APPLICANT: Cain, Scott
; APPLICANT: Rundlett, Stephen E.
; APPLICANT: Ramachandran, Rakesh
; TITLE OF INVENTION: NOVEL NUCLEIC ACID AND PROTEIN HOMOLOGS AND USES THEREOF
; FILE REFERENCE: APX-001CN
; CURRENT APPLICATION NUMBER: US/10/098,754
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 09/791539
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 21107
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19036
; LENGTH: 806
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

; NAME/KEY: misc_feature
; LOCATION: (1)...(806)
; OTHER INFORMATION: n = A,T,C or G
US-10-098-754-19036

```
Query Match   21.0%;  Score 79.4;  DB 39;  Length 806;
Best Local Similarity   98.8%;  Pred. No. 2.2e-13;
Matches   80;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy  262 GGTGATGGAGATGCCAATCCATGGAAATCAGGTGGCGCCAGCCTATGTTGGTGGTAGCTATAGCAG 321
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db  423 GGTGATGGAGATGCCAATCCATGGAATCAGGTGGCACAGCCTATGTTGGTAGCTATAGCAG 364

Qy  322 AAGTCTTCTTTGGCAAAGATTC 342
        |||||||||||||||||||||| 
Db  363 AAGTCTTCTTTGGCAAAGATTC 343
```

RESULT 13
US-09-685-045-1053
Sequence 1053, Application US/09685045
; GENERAL INFORMATION:
; APPLICANT: Nehls, Michael
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Novel Human Polynucleotides and
; TITLE OF INVENTION: Polypeptides Encoded Thereby
; FILE REFERENCE: LEX-0055-USA
; CURRENT APPLICATION NUMBER: US/09/685,045
; CURRENT FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 60/158,741
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 1508
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1053
; LENGTH: 125
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(125)
; OTHER INFORMATION: n = A,T,C or G
US-09-685-045-1053

```
Query Match   20.0%;  Score 75.6;  DB 27;  Length 125;
Best Local Similarity   83.1%;  Pred. No. 1.5e-12;
Matches   103;  Conservative   0;  Mismatches   19;  Indels   2;  Gaps   2;

Qy  198 CTTGGTTTTATGTCAACATGCCTTTGACTGT-TGAAAAGAGACCTTTGGCACACATTGAA 256
        |||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||||| 
Db    1 CTTGGTTGTATGCTCAACNTGCCTTTGCCTGTNTGAAAAGAGACCTTTGGCACACATTGAA 60

Qy  257 GGGATGGTGATGGAGATGCCAATCCATGGAAATCAGGTGGCGCCAGCCTATGTTGGTAGCTAT 316
        || |||||||| |||||||||||||||||  |||||||||||||||  |||||||||||||| 
Db   61 GGNATGGTGATNGNGATGCCNNTCCANGGAATCANGGGGGC-CNTCTAIGNCCCTTGCTCT 119

Qy  317 AGCA 320
        |||| 
Db  120 AGCA 123
```

RESULT 14
US-09-362-510-51216
Sequence 51216, Application US/09362510
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-759GCON1
; CURRENT APPLICATION NUMBER: US/09/362,510
; CURRENT FILING DATE: 1999-07-27
; EARLIER APPLICATION NUMBER: US 09/221,820
; EARLIER FILING DATE: 1998-12-30

```
;     NUMBER OF SEQ ID NOS: 62165
;     SOFTWARE: FastSEQ for Windows Version 3.0
;     SEQ ID NO 51216
;         LENGTH: 381
;         TYPE: DNA
;         ORGANISM: Homo sapiens
;         FEATURE:
;             NAME/KEY: misc_feature
;             LOCATION: (1)...(381)
;             OTHER INFORMATION: n = A,T,C or G
US-09-362-510-51216

Qy    326  CTTCTTGGCAAAAGATTCCTCCCCGGGGAAGGAAAGGTACCATTGGAGAACCATGC   377
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  CTTCTTGGCAAAAGATTCCTCCAGGGAAGGTAGGTACCATTGGAGAACCATGC   52

Query Match         12.9%;  Score 48.8;  DB 17;  Length 381;
Best Local Similarity  96.2%;  Pred. No. 0.00049;
Matches    50;  Conservative   0;  Mismatches   2;  Indels   0;  Gaps   0;

RESULT 15
US-09-362-510A-51216
;     Sequence 51216, Application US/09362510A
;     GENERAL INFORMATION:
;         APPLICANT:  Hyseq, Inc.
;         TITLE OF INVENTION:  NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;         TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;         FILE REFERENCE:  20411-7759CON1
;         CURRENT APPLICATION NUMBER:  US/09/362,510A
;         CURRENT FILING DATE:  1999-07-27
;         PRIOR APPLICATION NUMBER:  US 09/221,820
;         PRIOR FILING DATE:  1998-12-30
;         NUMBER OF SEQ ID NOS:  62165
;         SOFTWARE:  FastSEQ for Windows Version 3.0
;     SEQ ID NO 51216
;         LENGTH: 381
;         TYPE: DNA
;         ORGANISM: Homo sapiens
;         FEATURE:
;             NAME/KEY: misc_feature
;             LOCATION: (1)...(381)
;             OTHER INFORMATION: n = A,T,C or G
US-09-362-510A-51216

Qy    326  CTTCTTGGCAAAAGATTCCTCCCCGGGGAAGGAAAGGTACCATTGGAGAACCATGC   377
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  CTTCTTGGCAAAAGATTCCTCCAGGGAAGGTAGGTACCATTGGAGAACCATGC   52

Query Match         12.9%;  Score 48.8;  DB 17;  Length 381;
Best Local Similarity  96.2%;  Pred. No. 0.00049;
Matches    50;  Conservative   0;  Mismatches   2;  Indels   0;  Gaps   0;

Search completed: January 31, 2003, 02:17:22
Job time : 3495 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2003, 23:52:38 ; Search time 37 seconds
                 (without alignments)
                 4589.786 Million cell updates/sec

Title: US-09-803-719-222
Perfect score: 378
Sequence: 1 tacggctgcttaagacgact.........taccattggagaaccatgcg 378

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 396772 seqs, 224632407 residues

Total number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39.2	10.4	1408	9	US-09-938-842A-3096	Sequence 3096, Ap
2	31.6	8.4	2966	10	US-09-737-178-103	Sequence 103, App
3	31.6	8.4	3011	10	US-09-737-178-112	Sequence 112, App
4	31.6	8.4	3701	9	US-09-286-488-10	Sequence 10, Appl
5	31.6	8.4	3701	10	US-09-737-178-10	Sequence 10, Appl
6	31.4	8.3	568	10	US-09-864-761-12790	Sequence 12790, A
7	30.8	8.1	631	9	US-10-002-344-49	Sequence 49, Appl
8	30.8	8.1	440	9	US-09-796-692-4354	Sequence 4354, Ap
9	30.6	8.1	4285	10	US-09-745-763-204	Sequence 204, App
10	30.6	8.1	26591	10	US-09-764-877-2678	Sequence 2678, App
11	30.4	8.0	2648	10	US-09-822-849A-246	Sequence 246, Appl
12	30	7.9	1655	10	US-09-764-847-1518	Sequence 1518, Ap
13	30	7.9	14759	10	US-09-952-360-1	Sequence 1, Appl
14	30	7.9	116592	10	US-09-818-512-3	Sequence 3, Appli
15	29.6	7.8	438	10	US-09-783-590-12445	Sequence 12445, A
16	29.6	7.8	900	9	US-10-076-754-3	Sequence 3, Appli
17	29.6	7.8	900	9	US-10-076-773-3	Sequence 3, Appli
18	29.6	7.8	900	12	US-10-067-615-3	Sequence 3, Appli
19	29.6	7.8	1102	9	US-10-028-072-299	Sequence 299, App
20	29.6	7.8	1113	12	US-10-044-090-150	Sequence 150, App
21	29.6	7.8	1180	9	US-09-992-964-2	Sequence 2, Appli
22	29.6	7.8	1180	9	US-09-992-964-4	Sequence 4, Appli
23	29.6	7.8	1180	10	US-09-887-879-2	Sequence 2, Appli
24	29.6	7.8	1180	10	US-09-887-879-4	Sequence 4, Appli
25	29.6	7.8	1392	10	US-09-826-212-1	Sequence 1, Appli
26	29.4	7.8	3868	12	US-10-044-090-492	Sequence 492, App
27	29.2	7.7	322	9	US-09-928-457-97	Sequence 97, Appl
28	29.2	7.7	427	10	US-09-764-869-1477	Sequence 1477, Ap
29	29.2	7.7	2172	10	US-09-801-574-81	Sequence 81, Appl
30	29.2	7.7	63000	10	US-09-780-172-18	Sequence 18, Appl
31	29	7.7	3184	10	US-09-954-456-497	Sequence 497, App
32	28.6	7.6	2049	10	US-09-822-830A-312	Sequence 312, App
33	28.6	7.6	2520	10	US-09-938-842A-448	Sequence 448, App
34	28.6	7.6	4760	10	US-09-070-927A-351	Sequence 351, App
35	28.6	7.6	5046	10	US-09-725-735A-13	Sequence 13, Appl
36	28.6	7.6	147309	10	US-09-742-312-3	Sequence 3, Appli
37	28.4	7.6	2058	9	US-10-008-016-3	Sequence 3, Appli
38	28.4	7.6	2846	9	US-10-008-016-1	Sequence 1, Appli
39	28.4	7.5	15071	10	US-09-938-082A-29	Sequence 29, Appl
40	28.4	7.5	335913	9	US-09-754-853A-2	Sequence 2, Appli
41	28.4	7.5	335913	9	US-09-754-853A-3	Sequence 3, Appli
42	28.2	7.5	16106	10	US-09-764-877-2322	Sequence 2322, Ap
43	28.2	7.5	33206	10	US-09-880-107-2380	Sequence 2380, Ap
44	28.2	7.5	397658	10	US-09-813-320-3	Sequence 3, Appli
45	28.2	7.5	397658	10	US-09-813-320-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-938-842A-3096
Sequence 3096, Application US/0993842A
Patent No. US2002016037BA1
GENERAL INFORMATION:
  APPLICANT: Harper, Jeff
  APPLICANT: Kreps, Joel
  APPLICANT: Wang, Xun
  APPLICANT: Zhu, Tong
  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAIN[]
  TITLE OF INVENTION: SAME, AND METHODS OF USE
  FILE REFERENCE: SCRIP1300-3
  CURRENT APPLICATION NUMBER: US/09/938,842A
  CURRENT FILING DATE: 2001-08-24
  PRIOR APPLICATION NUMBER: US 60/227,866
  PRIOR FILING DATE: 2000-08-24
  PRIOR APPLICATION NUMBER: US 60/264,647
  PRIOR FILING DATE: 2001-01-16
  PRIOR APPLICATION NUMBER: US 60/300,111
  PRIOR FILING DATE: 2001-06-22
  NUMBER OF SEQ ID NOS: 5379
  SEQ ID NO 3096
  LENGTH: 1408
  TYPE: DNA
  ORGANISM: Arabidopsis thaliana
US-09-938-842A-3096

Query Match 10.4%; Score 39.2; DB 9; Length 1408;
Best Local Similarity 53.2%;
Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 106 GGAGCTGAAGGATGCACGGTTGTTAAGCCCCTGTTCTTTCCGTTGTTTAAATCTAAATGTT 165
         |||| || ||| | ||| || || ||||| || | |||| || ||| |||| | |||
Db 362 GGAGACCATATATGGACTTCTGGAACTCCGCTTAACATATCTCGTGTGTTACACACATGCG 421

Qy 166 CTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTATGCTCAACATGCTTTGAC 225
         |||||| | ||| | ||| || ||| | || ||| |||||| |||| |||||| |||||
Db 422 CATTTAATTATATAAAATCTTATTGACACAAAATTCTTGTTTTAACGTTTTTTTTGTA 481

Qy 226 TGTTGAAAGAGAGACCTTTGGCACACATTGAAGGGAT 261
         ||| | | || ||| || || || | |||
Db 482 AGTAAAGTAAAGCTTTATTATTAAAACAAAGAACAAAT 517

RESULT 2
US-09-737-178-103
Sequence 103, Application US/09737178
Patent No. US2001029295A1
GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C9
; CURRENT APPLICATION NUMBER: US/09/737,178
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 103
; LENGTH: 2966
; TYPE: DNA
; ORGANISM: Babesia microti
US-09-737-178-103 :

Query Match　　　　　8.4%;　Score 31.6;　DB 10;　Length 2966;
Best Local Similarity　51.4%;　Pred. No. 4.4;
Matches　73;　Conservative　0;　Mismatches　69;　Indels　0;　Gaps　0;

Qy　142　TTTTCCGTTGTTTAATCTAAATGTTCTTTGGAATAAAAAAAACCTCCCTGCCAAGTAGTACTTG　201
　　　　　　||||　||||　||||||　|||||　||||||||　|||　|||　|　||||　|　|||||
Db　271　TATTCTACTTTTAATATAATAATTTTTTATTCAATAATATACTCTTCAATATTTTGTACTAT　330

Qy　202　GTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGAT　261
　　　　　　||　|||　||||　|||　||||||||||||||　|||　|||||　||||||　||　||||||
Db　331　TTTATAATCATATATATATATAATAATATACTCTTCAATATTTTGATAATATATCAAATAAT　390

Qy　262　GGTGATGGAGATGCCAATCCAT　283
　　　　　　||||||　||　||||　|||||||
Db　391　GATGATATACATGAATATGCAT　412

RESULT 3
US-09-737-178-112
Sequence 112, Application US/09737178
Patent No. US2001029295A1
GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C9
; CURRENT APPLICATION NUMBER: US/09/737,178
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 112
; LENGTH: 3011
; TYPE: DNA
; ORGANISM: Babesia microti
US-09-737-178-112

Query Match　　　　　8.4%;　Score 31.6;　DB 10;　Length 3011;
Best Local Similarity　51.4%;　Pred. No. 4.4;
Matches　73;　Conservative　0;　Mismatches　69;　Indels　0;　Gaps　0;

RESULT 4
US-09-286-488-10
Sequence 10, Application US/09286488
Patent No. US2002016913GA1
GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C3
; CURRENT APPLICATION NUMBER: US/09/286,488
; CURRENT FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 3701
; TYPE: DNA
; ORGANISM: Babesia microti
US-09-286-488-10

Query Match　　　　　8.4%;　Score 31.6;　DB 9;　Length 3701;
Best Local Similarity　51.4%;　Pred. No. 4.8;
Matches　73;　Conservative　0;　Mismatches　69;　Indels　0;　Gaps　0;

Qy　142　TTTTCCGTTGTTTAATCTAAATGTTCTTTGGAATAAAAAAAACCTCCCTGCCAAGTAGTACTTG　201
　　　　　　||||　||||　||||||　|||||　||||||||　|||　|||　|　||||　|　|||||
Db　271　TATTCTACTTTTAATATAATAATTTTTTATTCAATAATATACTCTTCAATATTTTGTACTAT　330

Qy　202　GTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGAT　261
　　　　　　||　|||　||||　|||　||||||||||||||　|||　|||||　||||||　||　||||||
Db　331　TTTATAATCATATATATATATAATAATATACTCTTCAATATTTTGATAATATATCAAATAAT　390

Qy　262　GGTGATGGAGATGCCAATCCAT　283
　　　　　　||||||　||　||||　|||||||
Db　391　GATGATATACATGAATATGCAT　412

RESULT 5
US-09-737-178-10
Sequence 10, Application US/09737178
Patent No. US2001029295A1
GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C9
; CURRENT APPLICATION NUMBER: US/09/737,178
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10

Qy　142　TTTTCCGTTGTTTAATCTAAATGTTCTTTGGAATAAAAAAAACCTCCCTGCCAAGTAGTACTTG　201
　　　　　　||||　||||　||||||　|||||　||||||||　|||　|||　|　||||　|　|||||
Db　999　TATTCTACTTTTAATATAATAATTTTTTATTCAATAATATACTCTTCAATATTTTGTACTAT　1058

Qy　202　GTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGAT　261
　　　　　　||　|||　||||　|||　||||||||||||||　|||　|||||　||||||　||　||||||
Db　1059　TTTATAATCATATATATATATAATAATATACTCTTCAATATTTTGATAATATATCAAATAAT　1118

Qy　262　GGTGATGGAGATGCCAATCCAT　283
　　　　　　||||||　||　||||　|||||||
Db　1119　GATGATATACATGAATATGCAT　1140

; LENGTH: 3701
; TYPE: DNA
; ORGANISM: Babesia microti
US-09-737-178-10

Qy   Best Local Similarity   51.4%;   Score 31.6;   DB 10;   Length 3701;
Matches   73;   Conservative   0;   Mismatches   4,8;   Indels   0;   Gaps   0;

Qy   142 TTTTCCGTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTACTTG 201
         ||||| |||||||||| ||| || ||||| ||||| ||||||||| | ||||||
Db   999 TATTCCTACTTTTTAATATATAATTTTTTTTATTCCAATATACTCTTTCATATATTTGTACTTAT 1058

Qy   202 GTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAAGACCTTTGGCACACATTGAAGGGAT 261
         ||||| |||| |||||||||| |||||||||||||||||| ||||| |||| ||
Db 1059 TTTATATAATCATTATATATATTATAATTATATATTTGATTAATTGAATATATCAATAAT 1118

Qy   262 GGTGATGGAGATGCCAATCCAT 283
         | |||| ||| || || |||
Db 1119 GATGAATATACATGAATATGCAT 1140

RESULT 6
US-09-864-761-12790/c
; Sequence 12790, Application US/09864761
; Patent No. US2002004876341
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117

; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; FEATURE:
; SEQ ID NO 12790
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Homo sapiens
; OTHER INFORMATION: MAP TO AC011260.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.81
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.94
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.99
US-09-864-761-12790

Qy   Best Local Similarity   64.4%;   Score 31.4;   DB 10;   Length 568;
Matches   47;   Conservative   0;   Mismatches   26;   Indels   0;   Gaps   0;

Qy   93 GAATGTGTGTTCTGGAGCTGAAGGATGCACGGTTGTTGTTAAGCCCTGTTGTTAATCTAATGTT 152
         |||||||||| |||||||||||||| | || ||| | ||||||||||||| ||||||
Db   272 GAAATGTGTGTTCTCCAGCAGGATGAGTGAAAGGTTGTAAATATCACTTTTAGGTGTCTATCTAGT 213

Qy   153 TTAATCTAATGTT 165
         |||||||||| ||
Db   212 CTAATGTATAGTT 200

RESULT 7
US-10-002-344A-49/c
; Sequence 49, Application US/10002344A
; Patent No. US20020172959A1
; GENERAL INFORMATION:
; APPLICANT: Recipon, Herve
; APPLICANT: Sun, Yongming
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and
; FILE REFERENCE: DEX-0241
; CURRENT APPLICATION NUMBER: US/10/002,344A
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/242,998
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 49
; LENGTH: 631
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-002-344A-49

Qy   Best Local Similarity   47;   Score 30.8;   DB 9;   Length 631;
Matches   47;   Conservative   0;   Mismatches   27;   Indels   0;   Gaps   0;

Qy   106 GGAGCTGAAGGATGCACGGTTGTTGTTAAGCCCCTGTTCTTTCCGTTGTTAATCTAATGTT 165
         | |||||||| || || ||||||| | ||||||| ||||||| ||||| |||||||||||
Db   623 GCAACTGAAGGGCTAACATTTATTGAGCCCTCCTTCTTTTTGCCATGATCTTTGATATGTG 564

Qy   166 CTTTGGAATAAAA 179
         |||| ||||||||
Db   563 CTTTAAATTAAAAA 550

RESULT 8
US-09-796-692-4354/c
; Sequence 4354, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Galger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
FILE REFERENCE: 2077_001200
CURRENT APPLICATION NUMBER: US/09/796,692
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: 60/223,378
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 9597
SOFTWARE: FastSEQ for Windows Version 3.0

SEQ ID NO 4354
LENGTH: 440
TYPE: DNA
ORGANISM: Homo sapiens
US-09-796-692-4354

Query Match 93%; Conservative 8.1%; Score 30.6; DB 9; Length 440;
Best Local Similarity 47.2%; Pred. No. 3.9;
Matches 93; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

```
Qy  114 AGGATGCACGGTTGTTAAGCCCCCTGCTTCTTTCCGTTGTTTAATCTAAGTTCTTTGGAA 173
            |||  |||||||||||  |||||||||||| ||||||||||||||||||| |||||||
Db  395 ATGCTGCGTTGTTGTTCAGCAGCCCCTGTGTTCTGTGTGATTTGTTTTATTTTTCCTTTTT 336

Qy  174 TAAAAACCTCCCTGCCCAAGTAGTACTTGGTTTTATGCTCAAACATGCTTTGACTGTTGAAA 233
            ||| |||   |||  ||||||||||||||||||  ||||   |||| ||||||||||||
Db  335 TTTACATATATATGCAGGGAAGTAATTGGTACTGGTAGTGTATGTTGTTTCTATGTGGTTCAA 276

Qy  234 AGAGACCTTTGGCACACATTGAAGAGGATGGGTGATGGAGAGATGCCAAATCCATGGAATCAGGT 293
            ||| |||||||| ||||  ||||||||||||||  ||||||||||||||||||||||||||
Db  275 ATATGAATTTCGAACACACCAAGCCGCTAATGAGATAGCAGCTTTTTCTGGGACCCAGA 216

Qy  294 GGCGCAGCTATGTTGGT 310
            |||||||||||||
Db  215 GTCACAACCAAATTGAT 199
```

RESULT 9
US-09-745-763-204
Sequence 204, Application US/0974563
Patent No. US20020653941A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
         McCoy, John M.
         LaValle, Edward R.
         Collins-Racie, Lisa A.
         Evans, Cheryl
         Merberg, David
         Treacy, Maurice
         Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

NUMBER OF SEQUENCES: 219
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/745,763
FILING DATE: 18-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 204:
SEQUENCE CHARACTERISTICS:
LENGTH: 4285 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 204:
US-09-745-763-204

Query Match 93%; Conservative 8.1%; Score 30.6; DB 10; Length 4285;
Best Local Similarity 47.2%; Pred. No. 11;
Matches 93; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

```
Qy   114 AGGATGCACGGTTGTTAAGCCCCCTGCTTCTTTCCGTTGTTTAATCTAAGTTCTTTGGAA 173
             |||  |||||||||| ||||||||||||||||||||||||||||||||| |||||||
Db  2938 ATCCTGCGTTGTTGTTCAGCAGCCCCTGTGTTCTGTGTGATTTGTTTTATTTTTCCTTTTT 2997

Qy   174 TAAAAACCTCCCTGCCCAAGTAGTACTTGGTTTTATGCTCAAACATGCTTTGACTGTTGAAA 233
             ||| |||   |||  ||||||||||||||||||  ||||   |||| ||||||||||||
Db  2998 TTTACATATATATGCAGGGAAGTAATTGGTACTGGTAGTGTATGTTGTTTCTATGTGGTTCAA 3057

Qy   234 AGAGACCTTTGGCACACATTGAAGAGGATGGGTGATGGAGAGATGCCAAATCCATGGAATCAGGT 293
             ||| |||||||| ||||  ||||||||||||||  ||||||||||||||||||||||||||
Db  3058 ATATGAATTTCGAACACACCAAGCCGCTAATGAGATAGCAGCTTTTTCTGGGACCCAGA 3117

Qy   294 GGCGCAGCTATGTTGGT 310
             |||||||||||||
Db  3118 GTCACAACCAAATTGAT 3134
```

RESULT 10
US-09-764-877-2678/c
Sequence 2678, Application US/09764877
Patent No. US20020147140A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC005
CURRENT APPLICATION NUMBER: US/09/764,877
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2678
LENGTH: 26591
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-877-2678

Query Match   8.1%;   Score 30.6;   DB 10;   Length 26591;
Best Local Similarity   58.1%;
Matches 54; Conservative   0; Mismatches 39; Indels   0; Gaps   0;

Qy   145 TCCGTTCGTTTAATCTAATGTTCTTTGGAATAAAACCTCCCCCGCCAAGTAGTACTTGGTT 204
                ||||||| |||||| || || ||| | || |||||| |||||||| || ||
Db   15573 TCCGTTGCAAATTCAGTTCTTCCCTGCAGAAAAATTGTTCTTTTTAATTATTTTAATTGCTTT 15514

Qy   205 TTATGCTCAACATGCTTTGACTGTTGAAAAGAG 237
                |||||| ||||||| |||| ||| ||
Db   15513 TAATTTCACTACGGCTCTTACTGAATCAAAATAG 15481

RESULT 11
US-09-822-849A-246
; Sequence 246, Application US/0982849A
; Patent No. US2002045170A1
; GENERAL INFORMATION:
;   APPLICANT: Wong, Gordon G.
;   APPLICANT: Clark, Hilary
;   APPLICANT: Fechtel, Kim
;   APPLICANT: Agostino, Michael J.
;   APPLICANT: Howes, Steven H.
;   APPLICANT: Resnick, Richard J.
;   APPLICANT: Gulukota, Kamalakar
;   APPLICANT: Graham, James R.
;   APPLICANT: Genetics Institute, Inc.
;   TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
;   FILE REFERENCE: GIN 6403
;   CURRENT APPLICATION NUMBER: US/09/822,849A
;   CURRENT FILING DATE: 2001-09-04
;   PRIOR APPLICATION NUMBER: 60/195,582
;   PRIOR FILING DATE: 2000-04-06
;   NUMBER OF SEQ ID NOS: 598
;   SOFTWARE: Patentin Ver. 2.0
;
;   SEQ ID NO 246
;   LENGTH: 2648
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-822-849A-246

Query Match   8.0%;   Score 30.4;   DB 10;   Length 2648;
Best Local Similarity   51.5%;
Matches 70; Conservative   0; Mismatches 66; Indels   0; Gaps   0;

Qy   88 AGCTTGAATGTGTGTTCTCTGGAGCTGAAGGATGCACGGTTGTTGTTAAGCCCCTTGTTCTTTCC 147
                || ||| |||| | | || | || ||| || |||| |||||| | || ||
Db   1060 AACTTAAAATGTCAAAAAGGGAGGAGCAGCCGGAGTAGCAAAATAAAATTATGTTCACC 1119

Qy   148 GTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCCCGCCAAGTAGTACTTGGTTTTTA 207
                || |||||| ||| | ||| | | | || |||||| | |
Db   1120 CTTATGTAATAAAATACATTATGTAATAAAATAGGCTATATATAATAAGGTACACATTCAGGCA 1179

Qy   208 TGCTCAACATGCTTTG 223
                |||| ||||||| ||
Db   1180 TTTAAAACAGTGTG 1195

RESULT 12
US-09-764-847-1518/c
; Sequence 1518, Application US/09764847
; Patent No. US2002013276A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PC009
;   CURRENT APPLICATION NUMBER: US/09/764,847
;   CURRENT FILING DATE: 2001-01-17
;   Prior application data removed - consult PALM or file wrapper
;   NUMBER OF SEQ ID NOS: 2003
;   SOFTWARE: Patentin Ver. 2.0
;
;   SEQ ID NO 1518
;   LENGTH: 1655

; TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-847-1518

Query Match   7.9%;   Score 30;   DB 10;   Length 1655;
Best Local Similarity   54.5%;
Matches 60; Conservative   0; Mismatches 50; Indels   0; Gaps   0;

Qy   80 CTCTCACAAGCTTGAATGTGTGTTCTCTGGAGCTGAAGGATGCACGGTTGTTGTTAAGCCCCTGT 139
                |||| | ||||| ||| || | || | || ||| || |||| || | ||
Db   1221 CTCTAAGAATAAAAAATTAGATCCACGGTCACTCAAAAGAGGGAAAACTGATTTTTTTGTCTTTT 1162

Qy   140 TCTTTTCCGTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCCTGCC 189
                ||||| |||||||| |||| ||| | ||| | | | |
Db   1161 TGTTTTCCTTTGTTTTCCTTCCCTAAATTTCCTTTGAAGAACATGTCACCTGTC 1112

RESULT 13
US-09-952-360-1/c
; Sequence 1, Application US/09952360
; Patent No. US2002072066A1
; GENERAL INFORMATION:
;   APPLICANT: Styrkarsdottir, Unnur
;   APPLICANT: Johannsdottir, Vala Drofn
;   TITLE OF INVENTION: Human Osteoporosis Gene
;   FILE REFERENCE: 2345.2009-001
;   CURRENT APPLICATION NUMBER: US/09/952,360
;   CURRENT FILING DATE: 2001-09-13
;   PRIOR APPLICATION NUMBER: 09/661,887
;   PRIOR FILING DATE: 2000-09-14
;   NUMBER OF SEQ ID NOS: 2
;   SOFTWARE: FastSEQ for Windows Version 4.0
;
;   SEQ ID NO 1
;   LENGTH: 14759
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3639)...(3984)
;   NAME/KEY: CDS
;   LOCATION: (11757)...(12601)
US-09-952-360-1

Query Match   7.9%;   Score 30;   DB 10;   Length 14759;
Best Local Similarity   49.4%;
Matches 78; Conservative   0; Mismatches 80; Indels   0; Gaps   0;

Qy   29 AATGACCCAGCCGCGCTCTTAGAGAGAAACAATAGGAAAAACACCCAGCCGGAGTCTCTCACAA 88
                ||| || ||| | || || |||| ||||| ||| ||||| | |||||| || | ||
Db   8179 AATGAAACAGTTGCAGTTCAGAGGGTTAATGTAACTTTCCTAAGGTCACAGAGCTAGCAA 8120

Qy   89 GCTTGAATGTGTGTTCTCTGGAGCTGAAGGATGCACGGTTGTTGTTAAGCCCCTTGTTCTTTTCCG 148
                || ||| |||| | || | | || || || | || || |||| || || |||
Db   8119 ATTAAAAGTAGATATATATAAATTGAATCTACTCGGCTTCTCAGGTTTGTTTTTTTTGTCTTT 8060

Qy   149 TTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCT 186
                ||||| ||| |||| | || | |||| | || |
Db   8059 TTGTTTTCTTTTCTTGTTTTTTGTGACACAGGGTCTCACT 8022

RESULT 14
US-09-818-512-3/c
; Sequence 3, Application US/09818512
; Patent No. US2002142416A1
; GENERAL INFORMATION:
;   APPLICANT: BEASLEY, Ellen et al.
;   TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
;   TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
;   TITLE OF INVENTION: THEREOF
;   FILE REFERENCE: CL001192
;   CURRENT APPLICATION NUMBER: US/09/818,512
;   CURRENT FILING DATE: 2001-03-28
;   NUMBER OF SEQ ID NOS: 4

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 116592
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: (1)...(116592)
;     OTHER INFORMATION: n = A,T,C or G
US-09-818-512-3

  Query Match      49.4%;  Score 30;  DB 10;  Length 116592;
  Best Local Similarity   7.9%;           Pred. No. 70;
  Matches   78;  Conservative   0;  Mismatches  80;  Indels   0;  Gaps   0;

Qy     25 GGGGAATGACGAGCGGCTCTTAGAGGAACATATGGAAAACACCCAAGCCGGAGTCTCTC 84
          |||  |  |||||       ||||  ||| | | |        | |   |||||||||
Db  20183 GGGAAGCACAGCAAAGGGCTCTGAGAATAAATTAAAATATAAACCTGTTGCCAAAGCTCCAA 20124

Qy     85 ACAAGCTTGAATGTGTCGTTCTGGAGCTGAAGGATGCACGGTTGTTAAGCCCCTGTTCTTT 144
          || | || ||   | | ||||| |||   |  ||||  |  |||| | ||   |  |  |
Db  20123 ACTAACTAATAATAAATGGTTCTGTGGATAGGAAAAGACCCAAAGTAGCCATAGCACAGTCTTTGA 20064

Qy    145 TCCGTTGTTTAATCTAATGTTCTTTGGAATAAAAAACCT 182
          |  | | | |  | | |  ||| |  ||  |   | |||
Db  20063 ACTGATACTGGAAATACTATTCACAGGTGAGAAAACTT 20026


RESULT 15
US-09-783-590-12445
Sequence 12445, Application US/09783590
Patent No. US20020110850A1
GENERAL INFORMATION:
  APPLICANT: Dillon, Patrick J.
  APPLICANT: Haseltine, William A.
  APPLICANT: Li, Haodong
  APPLICANT: Rosen, Craig A.
  APPLICANT: Ruben, Steven M.
  TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
  FILE REFERENCE: PO-16.2C1
  CURRENT APPLICATION NUMBER: US/09/783,590
  CURRENT FILING DATE: 2000-02-15
  PRIOR APPLICATION NUMBER: 08/420,856
  PRIOR FILING DATE: 1995-04-12
  PRIOR APPLICATION NUMBER: 08/346,731
  PRIOR FILING DATE: 1994-11-21
  NUMBER OF SEQ ID NOS: 12485
  SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12445
;   LENGTH: 438
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;     NAME/KEY: misc feature
;     LOCATION: (177)
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: misc feature
;     LOCATION: (276)
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: misc feature
;     LOCATION: (333)
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: misc feature
;     LOCATION: (334)
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: misc feature
;     LOCATION: (335)
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: misc feature
;     LOCATION: (345)
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: misc feature
;     LOCATION: (375)
```

```
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: misc feature
;     LOCATION: (395)
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: misc feature
;     LOCATION: (404)
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: misc feature
;     LOCATION: (407)
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: misc feature
;     LOCATION: (422)
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: misc feature
;     LOCATION: (429)
;     OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-12445

  Query Match      61.0%;  Score 29.6;  DB 10;  Length 438;
  Best Local Similarity   7.8%;           Pred. No. 8.1;
  Matches   47;  Conservative   0;  Mismatches  30;  Indels   0;  Gaps   0;

Qy     93 GAATGTGTTCTGGAGCTGAAGGATGCACGGTTGTTAAGCCCCTGTTCTTTCCGTTGT 152
          ||||| ||||||  | ||| ||||||    ||  | |||  |    | | |  | |
Db    150 GAATGGTGTTCTGCACCTGATGGATGNAAATGTTTGTAAAATGTCTGTTAGATCCATTTA 209

Qy    153 TTAATCTAATGTTCTTT 169
          |  | | |  | | |
Db    210 TTTATTTATTTGTTTGT 226


Search completed: January 31, 2003, 01:17:11
Job time : 157 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2003, 23:50:38 ; Search time 52 seconds
(without alignments)
2229.304 Million cell updates/sec

Title: US-09-803-719-222
Perfect score: 378
Sequence: 1 tacggctgcttaagacgact..........taccattggagaaccatgcg 378
Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0
Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
    1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
    2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
    3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
    4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
    5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
    6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match Length	DB	ID	Description
c	1	31.6	8.4	4	3701	US-08-845-258-10	Sequence 10, Appl
c	2	31.6	8.4	4	3701	US-08-990-571-10	Sequence 10, Appl
c	3	31.6	8.4	4	3701	US-08-723-142A-10	Sequence 10, Appl
c	4	31.6	8.0	4	3701	US-09-528-784A-10	Sequence 10, Appl
c	5	30.4	7.9	4	20598	US-09-593-995-10	Sequence 10, Appl
c	6	30	7.9	4	7003	US-09-134-001C-1029	Sequence 7, App
c	7	29.6	7.8	4	900	US-08-378-548-24	Sequence 24, Appl
c	8	29.6	7.8	4	1392	US-09-006-353A-1	Sequence 3, Appl
c	9	29.6	7.8	4	1392	US-09-573-986-1	Sequence 1, Appl
c	10	29.6	7.8	4	4000	US-08-687-289A-2	Sequence 2, Appl
c	11	29.4	7.8	3	3834	US-09-209-668-18	Sequence 18, Appl
c	12	29.4	7.8	4	3854	US-08-365-470-1	Sequence 1, Appl
c	13	29.4	7.8	2	3858	US-09-344-15C-98	Sequence 98, Appl
c	14	29.4	7.8	4	3863	US-09-009-490A-88	Sequence 88, Appl
c	15	29.4	7.8	4	3863	US-08-482-073-1	Sequence 1, Appl
c	16	29.4	7.8	6	3863	5217870-1	Patent No. 5217870
c	17	29.2	7.7	1	864	US-08-204-675-3	Sequence 3, Appl
c	18	29.2	7.7	4	864	US-08-660-754-3	Sequence 3, Appl
c	19	29.2	7.7	2	864	US-08-796-364-3	Sequence 3, Appl
c	20	29.2	7.7	5	864	PCT-US95-02520-3	Sequence 3, Appl
c	21	29.2	7.7	4	4933	US-08-204-675-2	Sequence 2, Appl
c	22	29.2	7.7	2	4933	US-08-660-754-2	Sequence 2, Appl
c	23	29.2	7.7	2	4933	US-08-796-364-2	Sequence 2, Appl
c	24	29.2	7.7	5	4933	PCT-US95-02520-2	Sequence 2, Appl
c	25	29			162450	US-09-345-882-1	Sequence 1, Appl
c	26	28.8	7.6	1	420	US-08-504-718D-13	Sequence 13, Appl
c	27	28.8	7.6	1	420	US-08-504-718D-14	Sequence 14, Appl
c	28	28.8	7.6	1	420	US-08-504-718D-15	Sequence 15, Appl
c	29	28.8	7.6	1	420	US-08-504-718D-16	Sequence 16, Appl
c	30	28.8	7.6	4	1205	US-09-017-754A-19	Sequence 19, Appl
c	31	28.4	7.5	4	2259	US-07-828-700-7	Sequence 7, Appl
c	32	28.4	7.5	4	2358	US-09-134-001C-1029	Sequence 1029, Ap
c	33	28.2	7.5	1	1068	US-08-427-640-7	Sequence 7, Appl
c	34	28	7.4	3	376	US-08-981-803-9	Sequence 9, Appl
c	35	28	7.4	3	376	US-08-981-803-23	Sequence 23, Appl
c	36	28	7.4	4	376	US-08-983-440-9	Sequence 9, Appl
c	37	28	7.4	4	376	US-09-221-017B-169	Sequence 169, Appl
c	38	28	7.4	4	699	US-08-983-440-23	Sequence 23, Appl
c	39	28	7.4	4	11478	US-09-803-440-29	Sequence 29, Appl
c	40	28	7.4	4	11478	US-08-981-803-29	Sequence 29, Appl
c	41	27.6	7.4	2	2298	US-08-951-648-1	Sequence 1, Appl
c	42	27.6	7.3	3	2298	US-09-174-437-1	Sequence 1, Appl
c	43	27.6	7.3	3	3195	US-08-951-648-5	Sequence 5, Appl
c	44	27.6	7.3	3	3195	US-09-174-437-5	Sequence 5, Appl
c	45	27.6	7.3	2	4389	US-08-951-648-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-845-258-10
Sequence 10, Application US/08845258
Patent No. 6183976
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
APPLICANT: Sleath, Paul R.
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/845,258
FILING DATE: 24-APR-1997
CLASSIFICATION: 435
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.426C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206)682-6031
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 3701 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-845-258-10

Query Match 8.4%; Score 31.6; DB 4; Length 3701;
Best Local Similarity 51.4%;
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```
Qy  142 TTTTCCGTTGTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTACTTG 201
          | ||| |||||| |||||||| ||| |||| | ||||| |||| | |||||
Db  999 TATTCTACTTTTTAATATAATTTTTTTATTCAATAATATACTCTTCATATATTTGTACTAT 1058
```

Qy   202 GTTTTATGCTCAACATGCTCTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGAT 261
          | || ||| ||| |||
Db   1059 TTTATATAATCATATATTATTATATTATATTTGATAATTGAATATATCAATAAT 1118

Qy   262 GGTGATGGAGATGCCAATCCAT 283
Db   1119 GATGAATATACATGAATATGCAT 1140

RESULT 2
US-08-990-571-10
Sequence 10, Application US/08990571
Patent No. 6214971
; GENERAL INFORMATION:
;   APPLICANT: Reed, Steven G. et al.
;   TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B. M
;   NUMBER OF SEQUENCES: 79
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: SEED AND BERRY
;   STREET: 6300 Columbia Center, 701 Fifth Avenue
;   CITY: Seattle
;   STATE: Washington
;   COUNTRY: USA
;   ZIP: 98104
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/990,571
;   FILING DATE: 11-DEC-1997
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Maki, David J.
;   REGISTRATION NUMBER: 31,392
;   REFERENCE/DOCKET NUMBER: 210121.426C2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (206) 622-4900
;   TELEFAX: (206)682-6031
; INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 3701 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-08-990-571-10

Query Match   8.4%; Score 31.6; DB 4; Length 3701;
Best Local Similarity   51.4%; Pred. No. 1.6;
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy   142 TTTTCCGTTGTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTG 201
          | ||| | ||||||| ||| ||| ||| ||||||||| |||| | | |||||
Db   999 TATTTCTACTTTTTAATTAATAATTTTTATTCAATAATAATACTCTTTCATATTTTTGTTGTACTAT 1058

Qy   202 GTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGAT 261
          | || || ||| |||
Db   1059 TTTATATAATCATATATTATTATATTATATTTGATAATTGAATATATCAATAAT 1118

Qy   262 GGTGATGGAGATGCCAATCCAT 283
Db   1119 GATGAATATACATGAATATGCAT 1140

RESULT 3
US-08-723-142A-10
Sequence 10, Application US/08723142A
Patent No. 6306396
; GENERAL INFORMATION:
;   APPLICANT: Reed, Steven G.
;   APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond
; APPLICANT: Sleath, Paul R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: SEED AND BERRY
;   STREET: 6300 Columbia Center, 701 Fifth Avenue
;   CITY: Seattle
;   STATE: Washington
;   COUNTRY: USA
;   ZIP: 98104
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/723,142A
;   FILING DATE: 01-OCT-1996
;   CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
;   NAME: Maki, David J.
;   REGISTRATION NUMBER: 31,392
;   REFERENCE/DOCKET NUMBER: 210121.426
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (206) 622-4900
;   TELEFAX: (206)682-6031
; INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 3701 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-08-723-142A-10

Query Match   8.4%; Score 31.6; DB 4; Length 3701;
Best Local Similarity   51.4%; Pred. No. 1.6;
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy   142 TTTTCCGTTGTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTG 201
          | ||| | ||||||| ||| ||| ||| ||||||||| |||| | | |||||
Db   999 TATTTCTACTTTTTAATTAATAATTTTTATTCAATAATAATACTCTTTCATATTTTTGTTGTACTAT 1058

Qy   202 GTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGAT 261
          | || || ||| |||
Db   1059 TTTATATAATCATATATTATTATATTATATTTGATAATTGAATATATCAATAAT 1118

Qy   262 GGTGATGGAGATGCCAATCCAT 283
Db   1119 GATGAATATACATGAATATGCAT 1140

RESULT 4
US-09-528-784A-10
Sequence 10, Application US/09528784A
Patent No. 6451315
; GENERAL INFORMATION:
;   APPLICANT: Reed, Steven G.
;   APPLICANT: Lodes, Michael J.
;   APPLICANT: Houghton, Raymond L.
;   APPLICANT: Sleath, Paul R.
;   APPLICANT: McNeill, Patricia D.
;   TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
;   TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
;   FILE REFERENCE: 210121.426C4
;   CURRENT APPLICATION NUMBER: US/09/528,784A
;   CURRENT FILING DATE: 2000-03-17
;   NUMBER OF SEQ ID NOS: 90
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10
;   LENGTH: 3701
;   TYPE: DNA

; ORGANISM: Babesia microti
us-09-528-784A-10

Qy   142 TTTTCCGTTGTTTAATCTAATGTTCTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTG 201
         |||||| ||||| ||||||||| || ||||||||||||||||||||||||||||||||
Db   999 TATTCTACTTTTTAATATAATAATTTTTTTATTCAATAAATTACTCTTTCATATTTTGTACTAT 1058

Qy   202 GTTTTATGTCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGAT 261
         || ||| || ||| | ||| ||||||| |||||||| ||||||||||| |||
Db 1059 TTTATATAATCATATATATATATATATATATATATATATATTGATAATTGAATATATCAATAAT 1118

Qy   262 GGTGATGGAGAATGCCAATCCAT 283
         || |||| ||||| ||||| |||
Db 1119 GATGATATACATGAATATGCAT 1140

; Query Match   73; Conservative   51.4%; Pred. No. 1.6;
; Best Local Similarity 8.4%;
; Matches   73; Conservative   51.4%;   Mismatches 69; Indels 0; Gaps 0;
;   Score 31.6; DB 4; Length 3701;

RESULT 5
US-09-593-995-10/c
; Sequence 10, Application US/09593995
; Patent No. 6406888
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Gao, Zeren
; TITLE OF INVENTION: HELICAL CYTOKINE ZALPHA33
; FILE REFERENCE: 99-38
; CURRENT APPLICATION NUMBER: US/09/593,995
; CURRENT FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: 60/139,121
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 20598
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-593-995-10

Qy   149 TTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTAT 208
         |||||| |||||||||||| ||||||||||||||||||||||||||||||||| ||||
Db 17425 TGGTTTTACTTTCTAAATGCATGGGACTCAATCGCAACCTTGTAAATACTACCTTGTAAATACTACC 17366

Qy   209 GCTCAACATGCTTTGACTGTTGAAAGAGACCTTTGCCACACATTGAAGGGATGGATG 268
         |||| |||||||||| |||||| ||||||||| |||||| ||||||| ||||
Db 17365 TTTCAATTTCATTTTACCATCACAAGAAAAGACGCGTTAAGCAAAATAAAACTCATTCTGATG 17306

; Query Match   64; Conservative   53.3%; Pred. No. 8.9;
; Best Local Similarity 8.0%;
; Matches   64; Conservative   53.3%;   Mismatches 56; Indels 0; Gaps 0;
;   Score 30.4; DB 4; Length 20598;

RESULT 6
US-08-378-548-24/c
; Sequence 24, Application US/08378548
; Patent No. 5981177
; GENERAL INFORMATION:
; APPLICANT: Demirjian, David C.
; APPLICANT: Casadaban, Malcolm J.
; APPLICANT: Weber, J. Mark
; APPLICANT: Gaines III, George L.
; TITLE OF INVENTION: Protein Fusion Method and Constructs
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, Ltd.
; STREET: Ten South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/378,548
; FILING DATE: 25-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 95,181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7003 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..7003
; OTHER INFORMATION: /note= "MueII315, Figure 14"
US-08-378-548-24

Qy   116 GATGGCACGGTTGTTAAGCCCCTGTTCTTTTCCGTTGTGTTTTAAATCTAAATGGTTCTTTGGAATA 175
         | |||| |||||| || ||| ||| ||| ||||||||| || ||| ||| ||||| ||||||||
Db 3509 GAAACAGGCTCGTTATCCCCGTTGTACTTATCATTGATAAAAAACTACTGATAGTTCCGGTA 3450

Qy   176 AAAACCTCCCTGCCAAGTAGTACTTGGTTTTATG 209
         ||| |||||||| |||||||| |||| |||||
Db 3449 GAAAGCCCGCTCTCAAGAGCAATTTGTTTAAAGG 3416

; Query Match   54; Conservative   57.4%; Pred. No. 7.2;
; Best Local Similarity 7.9%;
; Matches   54; Conservative   57.4%;   Mismatches 40; Indels 0; Gaps 0;
;   Score 30; DB 2; Length 7003;

RESULT 7
US-09-134-618-3/c
; Sequence 3, Application US/09134618
; Patent No. 6417328
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; TITLE OF INVENTION: NOVEL TRAIL RECEPTORS, NUCLEIC ACIDS ENCODING SAME, AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 480140.432
; CURRENT APPLICATION NUMBER: US/09/134,618
; CURRENT FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 900
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(897)
US-09-134-618-3

Qy   120 CACGGTTGTTAAGCCCCTGTTCTTTTTCCGTTGTTTTAAATCTAAATGTTTCTTTGGAATAAAAA 179
         |||||| ||||| |||| || ||||||||| ||| ||||| ||| ||||||||| |||
Db   429 CATGGTGCAGGAAACTTTATGTTTTTTGGATCGCATTTACAAAACTGTTACATGGGAAGCAAGA 370

Qy   180 CCTCCCTGCCAAGTAGTACTTGGTTTTATGTCTCAACATGCTTTGACTGTTGAAAAGAGAC 239
         |||||||| |||||||| ||||||||| ||||| ||||| |||| ||||||| |||||
Db   ... 239

; Query Match   83; Conservative   48.3%; Pred. No. 3.7;
; Best Local Similarity 7.8%;
; Matches   83; Conservative   48.3%;   Mismatches 89; Indels 0; Gaps 0;
;   Score 29.6; DB 4; Length 900;

```
Db    369 AGGTTCATTGTTGGAAGCGTTGGTGTTGGTGTAATCCACACCCTCTTGTTGCACGGGTTACAGGCTCC 310
Qy    240 CTTTGGCACACATTGAAGGGATGGTTGATGGAGATCCAACCCTTGTCACTCCCCCCTTGAAGCTG 310
Db    309 AGTATGTTCTGATCTATGAGATCCTGCTGGACACTCCTCCCCCTTGAAGCTG 258
```

RESULT 8

US-09-006-353A-1/c
; Sequence 1, Application US/0906353A
; Patent No. 6261801
; GENERAL INFORMATION:
APPLICANT: WEI, YING-FEI
APPLICANT: YU, GUO-LIANG
APPLICANT: GENTZ, REINER
APPLICANT: RUBEN, STEVEN
TITLE OF INVENTION: TUMOR NECROSIS FACTOR RECEPTOR 5
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: HUMAN GENOME SCIENCES, INC.
STREET: 9410 KEY WEST AVENUE
CITY: ROCKVILLE
STATE: MD
COUNTRY: US
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,353A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BROOKES, ANDERS A
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF341
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1392 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 183..959
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 183..260
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 261..959
US-09-006-353A-1

Query Match   7.8%;   Score 29.6; DB 4; Length 1392;
Best Local Similarity   48.3%;
Matches   83; Conservative   0; Mismatches   89; Indels   0; Gaps   0;

```
Qy    120 CACGGGTTGTTAAGCCCCTGTTGTTCTTTTCCGTTGTTTAATCTAATGTTGTTCTTTGGAATAAAA 179
          ||  ||||       |||||||||||||||||||||||||||   ||| |||||||||||||
Db    491 CATGGTGCCAGGAACTTTTATGTTGTTTTGATCTGTAAGTTGTTACAATGGGAAGCAAGA 432
Qy    180 CCTCCCTGCCAAGTAGTACTTGGTTTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGAC 239
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    431 AGGTTCATTGTTGGAAGCGTTGGTGTTGTAATCCACACCCTCTTGTTGCACGGGTTACAGGCTCC 372
Qy    240 CTTTGGCACACATTGAAGGGATGGTTGATGGAGATGCCAATCCATGGAATCAG 291
          ||| ||| |||||||||||||||||||||||||||||| ||||| |||||||
```

```
Db    371 AGTATGTTCTGATCTATGAGATCCTGCTGGACACTCCTCCCCCTTGAAGCTG 320
```

RESULT 9

US-09-573-986-1/c
; Sequence 1, Application US/09573986
; Patent No. 6455040
; GENERAL INFORMATION:
APPLICANT: WEI, YING-FEI
APPLICANT: NI, JIAN
APPLICANT: GENTZ, REINER
APPLICANT: RUBEN, STEVEN
TITLE OF INVENTION: TUMOR NECROSIS FACTOR RECEPTOR 5
FILE REFERENCE: 1488.1280004
CURRENT APPLICATION NUMBER: US/09/573,986
CURRENT FILING DATE: 2000-05-18
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 1392
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (183)..(959)
US-09-573-986-1

Query Match   7.8%;   Score 29.6; DB 4; Length 1392;
Best Local Similarity   48.3%;
Matches   83; Conservative   0; Mismatches   89; Indels   0; Gaps   0;

```
Qy    120 CACGGGTTGTTAAGCCCCTGTTGTTCTTTTCCGTTGTTTAATCTAATGTTGTTCTTTGGAATAAAA 179
          ||  ||||       |||||||||||||||||||||||||||   ||| |||||||||||||
Db    491 CATGGTGCCAGGAACTTTTATGTTGTTTTGATCTGTAAGTTGTTACAATGGGAAGCAAGA 432
Qy    180 CCTCCCTGCCAAGTAGTACTTGGTTTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGAC 239
          ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    431 AGGTTCATTGTTGGAAGCGTTGGTGTTGTAATCCACACCCTCTTGTTGCACGGGTTACAGGCTCC 372
Qy    240 CTTTGGCACACATTGAAGGGATGGTTGATGGAGATGCCAATCCATGGAATCAG 291
          ||| ||| ||||||||||||||||||||||||||||| |||||  |||||||
Db    371 AGTATGTTCTGATCTATGAGATCCTGCTGGACACTCCTCCCCCTTGAAGCTG 320
```

RESULT 10

US-08-687-289A-2
; Sequence 2, Application US/08687289A
; Patent No. 5981195
; GENERAL INFORMATION:
APPLICANT: Fuller, Forrest H.
APPLICANT: Krapcho, Karen J.
APPLICANT: Hammerland, Lance G.
TITLE OF INVENTION: CHIMERIC RECEPTORS AND METHODS FOR
TITLE OF INVENTION: IDENTIFYING COMPOUNDS ACTIVE AT
TITLE OF INVENTION: METABOTROPIC GLUTAMATE RECEPTORS AND
TITLE OF INVENTION: THE USE OF SUCH COMPOUNDS IN THE
TITLE OF INVENTION: TREATMENT OF NEUROLOGICAL DISORDERS
TITLE OF INVENTION: AND DISEASES
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSEQ for Windows 2.0

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,289A
; FILING DATE: July 25, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,526
; FILING DATE: July 26, 1995
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 32,327
; NAME: Warburg, Richard J.
; REFERENCE/DOCKET NUMBER: 220/004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4000 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-687-289A-2

Query Match   7.8%;   Score 29.6; DB 2; Length 4000;
Best Local Similarity   50.7%;
Matches   71; Conservative 0; Mismatches 7.5; Indels 0; Gaps 0;

Qy   35 GCAGCGGCTCTTAGAGGAACATATGGAAAACACCCAAGCCCGGAGTCTCTCACAAGCTTGA 94
Db 1738 GGAGTGTCCTGATGGGGAGTATAGTGAATGAGACAGAGTGCCCAGTTGCCTTGTTCTTTTCCGTTGTTTT 1797

Qy   95 ATGTGTGTTCTCGGAGCTGAAGGATGCACGGTTGTTAAGCCCCCTGTTCTTTTCCGTTGTTTT 154
Db 1798 AGATGACTTCTGGTCCAATGAGAACCACACCTCCTGCGAGCCCATTCCTGTCTCCGTTATCT 1857

Qy   155 AATCTAATGTTCTTTGGAAT 174
Db 1858 TGAGTGGAGTGACATAGAT 1877

RESULT 11
US-09-209-668-18
Sequence 18, Application US/09209668A
Patent No. 611517
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; APPLICANT: Xu, Xiaoxing S.
; TITLE OF INVENTION: METHODS OF MODULATING TUMOR NECROSIS FACTOR
; TITLE OF INVENTION: alpha-INDUCED EXPRESSION OF CELL ADHESION MOLECULES
; FILE REFERENCE: ISPH-0336
; CURRENT APPLICATION DATA:
; CURRENT APPLICATION NUMBER: US/09/209,668A
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 18
; LENGTH: 3834
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (117)..(1949)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: M24736/GenBank
; DATABASE ENTRY DATE: 1994-11-07
US-09-209-668-18

Query Match   7.8%;   Score 29.4; DB 3; Length 3834;
Best Local Similarity   49.1%;
Matches   78; Conservative 0; Mismatches 8.6; Indels 0; Gaps 0;

Qy   114 AGGATGCACGGGTTGTTAAGCCCCTGTTCTTTTCCGTTGTTTTAATCTAATGTTCTTTGGAA 173
Db 3192 AAGATGTCTAATAGTTATTCCCTATTTGTTTCTTCTCGTATGTTAGGGTGCTCTGGAAGA 3251

RESULT 12
US-08-365-470-1
Sequence 1, Application US/08365470
Patent No. 5632991
; GENERAL INFORMATION:
; APPLICANT: Gimbrone, Jr., Michael A.
; TITLE OF INVENTION: Antibodies Specific For E-selectin And The Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; SOFTWARE: Patentin Release #1.0, Version #1.25
; APPLICATION NUMBER: US/08/365,470
; FILING DATE: herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/102,510
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/850,802
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Markowicz, Karen R.
; REGISTRATION NUMBER: 36,351
; REFERENCE/DOCKET NUMBER: 0627.1350003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3854 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-365-470-1

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005

Query Match   7.8%;   Score 29.4; DB 1; Length 3854;
Best Local Similarity   49.1%;
Matches   78; Conservative 0; Mismatches 8.6; Indels 0; Gaps 0;

Qy   114 AGGATGCACGGGTTGTTAAGCCCCTGTTCTTTTCCGTTGTTTTAATCTAATGTTCTTTGGAA 173
Db 3192 AAGATGTCTAATAGTTATTCCCTATTTGTTTCTTCTCGTATGTTAGGGTGCTCTGGAAGA 3251

Qy   174 TAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTATTGCTCAACATGCTTTGACTGTTGAAA 233
Db 3252 GAGGAATGCCTGTGTGAGCAAGCACATTTGATGTTTGTTTTATTTAATGCAGAGATTTAACAATTCCAA 3311

Qy   234 AGAGACCTTTGGCACACATTGAAGGGATGGGTGATGGAGA 272
Db 3312 AGGAATCTCCAGTTTTCAGTTGATCACTGGCAATGAAAA 3350

RESULT 13
US-08-344-155C-98
Sequence 98, Application US/08341155C
Patent No. 5883082
GENERAL INFORMATION:
;   APPLICANT: Bennett and Stepkowski
;   TITLE OF INVENTION: Compositions and Methods for Preventing
;   TITLE OF INVENTION: and Treating Allograft Rejection
;   NUMBER OF SEQUENCES: 99
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Woodland Falls Corporate Park
;   STREET: 210 Lake Drive East, Suite 201
;   CITY: Cherry Hill
;   STATE: NJ
;   COUNTRY: USA
;   ZIP: 08002
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;   COMPUTER: IBM PS/2
;   OPERATING SYSTEM: PC-DOS
;   SOFTWARE: WORDPERFECT 5.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/344,155C
;   FILING DATE: No. 5883082ember 23, 1994
;   CLASSIFICATION: 514
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US91/05209
;   FILING DATE: July 23, 1991
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 939,855
;   FILING DATE: September 2, 1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/063,167
;   FILING DATE: 5/17/93
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/007,997
;   FILING DATE: 1/21/93
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 9/2/92
;   FILING DATE: 9/2/92
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/939,855
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/567,286
;   FILING DATE: 8/14/90
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Jane Massey Licata
;   REGISTRATION NUMBER: 32,257
;   REFERENCE/DOCKET NUMBER: ISPH-0098
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (609) 779-2400
;   TELEFAX: (609) 779-8488
;   INFORMATION FOR SEQ ID NO: 98:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 3858
;   TYPE: Nucleic Acid
;   STRANDEDNESS: Single
;   TOPOLOGY: Linear
;   ANTI-SENSE: No
US-08-344-155C-98

Query Match   7.8%;   Score 29.4;   DB 2;   Length 3858;
Best Local Similarity   49.1%;   Pred. No. 8.6;
Matches 78;   Conservative   0;   Mismatches 81;   Indels   0;   Gaps   0;

Qy   114 AGGATGCACGGTTGTTAAGCCCCTGTTGTTCTTCCGTTGTGTTTAATCTAATGTTCTTGGAA 173
          ||||| ||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   3216 AAGATGTCTAATACTTATTCCCTATTGTTGTTTCTTCTGTATGTTAGGGTGCTCTGGAAGA 3275

Qy   174 TAAAAACCTCCCTGCCAAGTAGTACTTGGTTTATGCTCAACATGCTTTGACTGTTGAAA 233
          | |||||||||||||||||||||||||||||| ||||| ||||||||||||||||| ||
Db   3276 GAGGAATGCCTGCTGTGTGAGCAAGCAATTTATGTTTATTTTATAAGCAGATTTAACAAT 3335

Qy   234 AGAGACCTTTGGCACACAATTGAAGGGATGGTGATTGGGAGA 272
          || ||| ||||| || ||||| ||||||| || || || || |
Db   3276 GAGGAATGCCTGCTGTGTGAGCAAGCAATTTATGTTTATTTTATAAGCAGATTTAACAATTCCAA 3335

Db   3336 AGGAATCTCCAGTTTTCAGTTGATCACTGGCAATGAAAA 3374

RESULT 14
US-09-009-490A-88
Sequence 88, Application US/09009490A
Patent No. 6300491
GENERAL INFORMATION:
;   APPLICANT: Bennett and Mirabelli
;   TITLE OF INVENTION: Oligonucleotide Modulation
;   TITLE OF INVENTION: of Cell Adhesion
;   NUMBER OF SEQUENCES: 95
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Law Office of Jane Massey Licata
;   STREET: 66 East Main Street
;   CITY: Marlton
;   STATE: NJ
;   COUNTRY: USA
;   ZIP: 08053
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;   COMPUTER: IBM PS/2
;   OPERATING SYSTEM: Windows 95
;   SOFTWARE: WORDPERFECT 6.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/009,490A
;   FILING DATE: January 20, 1998
;   CLASSIFICATION: 514
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 440,740
;   FILING DATE: May 12, 1995
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 063,167
;   FILING DATE: May 17, 1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 969,151
;   FILING DATE: February 10, 1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 007,997
;   FILING DATE: January 20, 1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 939,855
;   FILING DATE: September 2, 1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 567,286
;   FILING DATE: August 14, 1990
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Jane Massey Licata
;   REGISTRATION NUMBER: 32,257
;   REFERENCE/DOCKET NUMBER: ISPH-0268
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (609) 810-1515
;   TELEFAX: (609) 810-1454
;   INFORMATION FOR SEQ ID NO: 88:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 3858
;   TYPE: Nucleic Acid
;   STRANDEDNESS: Single
;   TOPOLOGY: Linear
;   ANTI-SENSE: no
US-09-009-490A-88

Query Match   7.8%;   Score 29.4;   DB 4;   Length 3858;
Best Local Similarity   49.1%;   Pred. No. 8.6;
Matches 78;   Conservative   0;   Mismatches 81;   Indels   0;   Gaps   0;

Qy   114 AGGATGCACGGTTGTTAAGCCCCTGTTGTTCTTCCGTTGTGTTTAATCTAATGTTCTTGGAA 173
          ||||| ||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   3216 AAGATGTCTAATAGTTATTCCCTATTGTTGTTTCTTCTGTATGTTAGGGTGCTCTGGAAGA 3275

Qy   174 TAAAAACCTCCCTGCCAAGTAGTACTTGGTTTATGCTCAACATGCTTTGACTGTTGAAA 233
          | |||||||||||||||||||||||||||||| ||||| ||||||||||||||||| ||
Db   3276 GAGGAATGCCTGCTGTGTGAGCAAGCAATTTATGTTTATTTTATAAGCAGATTTAACAATTCCAA 3335

```
Qy    234 AGAGACCTTTGGCACACATTGAAGGAGGATGGTGATGAGGAGA 272
          |||  ||  |||      ||  |||    |||  |  ||||
Db   3336 AGGAATCTCCAGTTTTCAGTTGATCACTGGCAATGAAA 3374
```

RESULT 15
US-08-482-073-1
; Sequence 1, Application US/08482073
; Patent No. 6307025
; GENERAL INFORMATION:
;   APPLICANT: Hession, Catherine A.
;   APPLICANT: Lobb, Roy R.
;   APPLICANT: Goelz, Susan E.
;   APPLICANT: Osborn, Laurelee
;   APPLICANT: Benjamin, Christopher D.
;   APPLICANT: Rosa, Margaret D.
;   TITLE OF INVENTION: ENDOTHELIAL CELL-LEUKOCYTE ADHESION
;   TITLE OF INVENTION: MOLECULES (ELAMS) AND MOLECULES INVOLVED IN LEUKOCYTE
;   TITLE OF INVENTION: ADHESION (MILAs)
;   NUMBER OF SEQUENCES: 25
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Neave
;   STREET: 1251 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   COUNTRY: United States of America
;   ZIP: 10020
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/482,073
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/08/486,336
;   FILING DATE:
;   APPLICATION NUMBER: US 07/608298
;   FILING DATE: 31-OCT-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: WO PCT/US 90/02357
;   FILING DATE: 27-APR-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/452675
;   FILING DATE: 18-DEC-1989
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/359516
;   FILING DATE: 01-JUN-1989
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/354151
;   FILING DATE: 28-APR-1989
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Haley Jr., James F.
;   REGISTRATION NUMBER: 27,794
;   REFERENCE/DOCKET NUMBER: B124CIP4
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 596-9000
;   TELEFAX: (212) 596-9090
;   TELEX: 14-8367
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 3863 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
US-08-482-073-1

Query Match   7.8%; Score 29.4; DB 4; Length 3863;
Best Local Similarity   49.1%; Pred. No. 8.6;
Matches   78; Conservative   0; Mismatches 81; Indels   0; Gaps   0;
```

```
Qy    114 AGGATGCACGGTTGTTGTTAAGCCCCTGTTCTTTTTCCGTTGTTGTTTAATCTAAATGTTTCTTTGGAA 173
          |   ||||          ||      |||    |||   | | |||||  ||  |  |||||| |  |  ||
Db   3275 GAGGAATGCCTGTGTGAGCAAGCAGCCATTTATGTTTATTTATTTATAAGCAGATTTAACAATTCCAA 3334

Qy    174 TAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTTTTATCTCAACATGCTTTTGACTGTTGAAA 233
          |  |   |||    | ||   |||  |  ||| | |||||    ||| ||| ||  || ||||  ||
Db   3215 AAGATGTCTAATAGTTATTCCCTATTTGTTTTCTTCTTGGTATGTTAGGGTGGTCTCTGGAAGA 3274

Qy    234 AGAGACCTTTGGCACACATTGAAGGAGGATGGTGATGAGGAGA 272
          |||  ||  |||      ||  |||    |||  |  ||||    | | ||
Db   3335 AGGAATCTCCAGTTTTCAGTTGATCACTGGCAATGAAA 3373
```

Search completed: January 31, 2003, 01:14:56
Job time : 88 secs

```
GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:            January 31, 2003, 01:17:19 ; Search time 2405 Seconds
                   (without alignments)
                   2545.486 Million cell updates/sec

Title:             US-09-803-719-222
Perfect score:     378
Sequence:          1 tacggctgcttaagacgact...........taccattggagaaccatgcg 378

Scoring table:     IDENTITY_NUC
                   Gapop 10.0 , Gapext 1.0
Searched:          16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters:    32308132

Post-processing: Minimum Match 0%
                 Maximum Match 100%
Minimum DB seq length: 0                    Listing first 45 summaries
Maximum DB seq length: 2000000000

Database :
        1: em_estba:*
        2: em_esthum:*
        3: em_estin:*
        4: em_estmu:*
        5: em_estov:*
        6: em_estpl:*
        7: em_estro:*
        8: em_htc:*
        9: gb_est1:*
       10: gb_est2:*
       11: gb_htc:*
       12: gb_est3:*
       13: gb_est4:*
       14: gb_est5:*
       15: em_estfun:*
       16: em_estom:*
       17: gb_gss:*
       18: em_gss_hum:*
       19: em_gss_inv:*
       20: em_gss_pln:*
       21: em_gss_vrt:*
       22: em_gss_fun:*
       23: em_gss_mam:*
       24: em_gss_mus:*
       25: em_gss_other:*
       26: em_gss_pro:*
       27: em_gss_rod:*
```

SUMMARIES

```
                         %
Result           Query
No.    Score  Match Length  DB  ID          Description
-----  -----  ------------  --  ----------  -----------
 c  1  167.8   44.4   451   14  BM987874    BM987874 UI-H-COO-
 c  2  118.8   31.4   439    9  AI455918    AI455918 ar81h09.x
 c  3  108.8    9     386    9  AA743908    AA743908 ob05b10.s
 c  4   79.4   21.0   806   12  BG197212    BG197212 RST16451
 c  5   62.6   16.6   349   17  AQ035618    AQ035618 CIT-HSP-2
 c  6   41.4   11.0   237   17  AQ035738    AQ035738 CIT-HSP-2
```

```
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

```
                                                      Pred.
        7   39.8   10.5   554   14  B0655384    B0655384 NXRV093_F
        8   39.2   10.4   439   17  BH853295    BH853295 SALK_0763
 c      9   38.8   10.3  1237   13  BI454874    BI454874 603173414
 c     10   38     10.1   585   14  BQ392762    BQ392762 NISC_mq26
 c     11   37.4    9.9   641    9  AL628007    AL628007 AL628007
 c     12   38     10.1   592   17  AQ922388    AQ922388 RPCI-23-2
 c     13   37      9.8   699   17  AZ108808    AZ108808 RPCI-23-2
 c     14   37      9.8   829   17  AZ191123    AZ191123 SP_1018_B
 c     15   36.8    9.7   473    9  AI449346    AI449346 mt07d11.x
 c     16   36.4    9.6   550   17  AQ775948    AQ775948 HS_2151_B
 c     17   36.4    9.6   798    9  AI350894    AI350894 qt13d12.x
 c     18   36      9.5   254    9  AI600177    AI600177 EST251880
 c     19   35.2    9.3   601   10  AW704405    AW704405 sk30f07.y
 c     20   35      9.3   807   17  CNS00AMP    CNS00AMP Drosophil
 c     21   34.8    9.2   398    9  AI781854    AI781854 EST262733
 c     22   34.8    9.2   554   17  AZ720218    AZ720218 RPCI-24-1
 c     23   34.8    9.2   695   14  BM973205    BM973205 UI-CF-EC1
 c     24   34.6    9.2   246   10  BB566524    BB566524 BB566524
 c     25   34.6    9.2   532   10  AZ182814    AZ182814 SP_0190_A
 c     26   34.6    9.2   935   13  BI861604    BI861604 603390466
 c     27   34.4    9.1   416   10  AV413421    AV413421 AV413421
 c     28   34.4    9.1   728   17  AZ400532    AZ400532 1M016GD23
 c     29   34.4    9.1   819   17  CNS0040P    CNS0040P Tetraodon
 c     30   34.2    9.0   562   10  AW231608    AW231608 687067H09
 c     31   34      9.0   649   10  BB481722    BB481722 BB481722
 c     32   33.8    8.9   571    9  AI828325    AI828325 wk83d06.x
 c     33   33.8    8.9   584   17  AQ419832    AQ419832 RPCI-11-1
 c     34   33.8    8.9   643   10  AW190856    AW190856 x142f09.x
 c     35   33.8    8.9   654   17  AZ083843    AZ083843 RPCI-23-2
 c     36   33.6    8.9   345   10  AW444676    AW444676 UI-H-BI3-
 c     37   33.6    8.9   491   14  BQ024523    BQ024523 UI-1-BB1p
 c     38   33.4    8.9   521   17  AZ182824    AZ182824 SP_0190_A
 c     39   33.4    8.8   247   10  BE067501    BE067501 MR4-BP035
 c     40   33.4    8.8   286    9  AI651773    AI651773 wb50b04.x
 c     41   33.4    8.8   337    9  AA711119    AA711119 vt56b05.r
 c     42   33.4    8.8   499   10  AW635726    AW635726 b137d10.w
 c     43   33.4    8.8   685   13  BJ068145    BJ068145 BJ068145
 c     44   33.4    8.8   899   17  AZ534719    AZ534719 ENTBW93TR
 c     45   33.2    8.8   488   12  BG136484    BG136484 EST477022
```

ALIGNMENTS

```
RESULT 1
BM987874/c
LOCUS       BM987874                451 bp    mRNA    linear    EST 17-JUN-2002
DEFINITION  UI-H-COO-asz-a-11-0-UI.s1 NCI_CGAP_Sub9 Homo sapiens cDNA clone
            IMAGE:5860149 3', mRNA sequence.
ACCESSION   BM987874
VERSION     BM987874.1        GI:19707263
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 451)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Seq primer: M13 FORWARD
            POLYA=yes.
FEATURES             Location/Qualifiers
     source          1 .451
```

```
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:5860149"
                    /clone_lib="NCI_CGAP_Sub9"
                    /tissue_type="mixed"
                    /dev_stage="mixed"
                    /lab_host="DH10B (Life Technologies)"
                    /note="Vector: pT7T3-Pac (Pharmacia) with a modified
                    polylinker; Site_1: EcoR I; Site_2: Not I; tissues:
                    Cholonic mucosa with Crohns disease, Cholonic mucosa with
                    ulcerative colitis, Fetal thymus, Cervix, Cervical
                    adenosquamous carcinoma, Ligament cells, Prostate
                    carcinoma, Bladder carcinoma, Brain oligodenroga ;
                    NCI_CGAP_Sub9 is a subtracted cDNA library constructed
                    according to Bonaldo, Lennon and Soares, Genome Research,
                    6:791-806, 1996. First strand cDNA synthesis was primed
                    with an oligo-dT primer containing a Not I site. Double
                    stranded cDNA was ligated to an EcoR I adaptor, digested
                    with Not I, and cloned directionally into pT7T3-Pac
                    vector. The oligonucleotide used to prime the synthesis of
                    first-strand cDNA contains a library tag sequence that is
                    located between the Not I site and the (dt)18 tail. The
                    sequence tags for this library are CGTC, AACG, GGGCC,
                    GGAAG, TAGC, TAAGC, ATGG, AGACA, ATCAC. For additional
                    Information, contact: Bento Soares, bento-soares@uiowa.edu

BASE COUNT      109 a       93 c       83 g       166 t
ORIGIN
                    TAG_LIB=UI-H-C00
                    TAG_TISSUE=Prostate Carcinoma
                    TAG_SEQ=ATGG"

Query Match       44.4%;   Score 167.8;   DB 14;   Length 451;
Best Local Similarity   98.8%;   Pred. No. 4.6e-38;
Matches  169;   Conservative   0;   Mismatches  2;   Indels  0;   Gaps  0;

Qy  207  ATGCTCAACACATGCTTTGACTGTGTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGA  266
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  451  ATGCTCAACACATGCTTTGACTGTGTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGA  392

Qy  267  TGGAGATGCCAATCCATGGAATCAGGTGTGGCGCGAGCTATGTTGGTAGCGTATAGCAGAAGTC  326
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  391  TGGAGATGCCAATCCATGGAATCAGGTGTGGCGCGAGCTATGTTGGTAGCGTATAGCAGAAGTC  332

Qy  327  TTCTTGGCCAAAGATTCCTCCGGGAAGGAAAGGTACCCAGCCTATGTTGGTAGCTA  377
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  331  TTCTTGGCCAAAGATTCCTCCAGGGAAGGAAAGGTACCACATTGGAGAACCATGC  281
```

```
FEATURES             Location/Qualifiers
     source          1..439
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2151713"
                     /clone_lib="Barstead colon HPLRB7"
                     /sex="male"
                     /dev_stage="adult, age 25"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAAGTCGAGGTGGAGACGGCGGCGCCCCTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [5' AATTCACTAGTAAT 3' and 5' ATTACTAGTG 3'], digested
                     with Not I and cloned into the Not I and Eco RI sites of
                     the modified pT7T3 vector. Library constructed by Bob
                     Barstead."

BASE COUNT       115 a       89 c       81 g       154 t
ORIGIN

Query Match       31.4%;   Score 118.8;   DB 9;   Length 439;
Best Local Similarity   98.4%;   Pred. No. 8e-24;
Matches  120;   Conservative   0;   Mismatches  2;   Indels  0;   Gaps  0;

Qy  256  AGGGATGGTGATGGAGATGCCAATCCATGGAATCAGGTGGCGCGCAGCCTATGTTGGTAGCTA  315
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  439  AGGGATGGTGATGGAGATGCCAATCCATGGAATCAGGTGGCACAGCCTATGTTGGTAGCTA  380

Qy  316  TAGCAGAAGTCTTCTTGGCCAAAGATTCCTCCCGGGAAGGAAAGGTACCCAGCCTATGTTGGT  375
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  379  TAGCAGAAGTCTTCTTGGCCAAAGATTCCTCCAGGGAAGGAAAGGTACCACATTGGAGAACCAT  320
```

```
RESULT 2
LOCUS       AI459918                  439 bp    mRNA    linear    EST 09-MAR-1999
AI459918/c
DEFINITION  ar81h09.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone
            IMAGE:2151713 3', mRNA sequence.
ACCESSION   AI459918
VERSION     AI459918.1   GI:4312799
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Contact: Wilson RK
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
FEATURES             Location/Qualifiers
     source
```

```
RESULT 3
LOCUS       AA743908                  386 bp    mRNA    linear    EST 19-FEB-1998
AA743908/c
DEFINITION  ob05b10.s1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1322279 3',
            mRNA sequence.
ACCESSION   AA743908
VERSION     AA743908.1   GI:2784658
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            unknown library type
            Insert Length: 508    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 234.
FEATURES             Location/Qualifiers
     source          1..386
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1322279"
                     /clone_lib="NCI_CGAP_Kid3"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
```

a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to a Not I and Eco RI adaptors
(Pharmacia), digested with Not I and Eco RI sites of the modified pT7T3 vector. mRNA
I and Eco RI sites of the modified pT7T3 vector. mRNA
source: 2 pooled kidneys. Library constructed by Bento Soares and
of normalization. Library went through one round
M. Fatima Bonaldo."

```
BASE COUNT          93 a         77 c         73 g        143 t
ORIGIN
```

Qy   266 ATGGAGAATGCCAATCCATGGAATCAGGTGGCGCCCAGCTATGTTGGTAGCTATAGCAGAAGT 325
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   386 ATGGAGAATGCCAATCCATGGAATCAGGTGGCACAGCTATGTTGGTAGCTATAGCAGAGT 327
Qy   326 CTTCTTGGCAAAGATTCCTCCCGGGAAGGAGAACGGTACCATTGGAGAACCATGC 377
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   326 CTTCTTGGCAAAGATTCCTCCAGGGAGGAGAAGGTACCATTGGAGAACCATGC 275

```
RESULT 4
BG197212/c
LOCUS          BG197212               806 bp    mRNA    linear    EST 21-APR-2001
DEFINITION     RST16451 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION      BG197212
VERSION        BG197212.1      GI:13718899
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1  (bases 1 to 806)
AUTHORS        Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
               Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.
               , Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith
               ,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
               ,J., Danzig,J. and Ducar,M.
TITLE          Creation of genome-wide protein expression libraries using random
               activation of gene expression
JOURNAL        Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE        21227151
COMMENT        Contact: Scott J. Cain
               Athersys, Inc.
               3201 Carnegie Ave, Cleveland, OH 44115, USA
               Tel: 216 431 9900
               Fax: 216 361 9596
               Email: scain@athersys.com
               High quality sequence stop: 417.
FEATURES             Location/Qualifiers
     source          1..806
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Athersys RAGE Library"
                     /cell_line="HT1080"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."
BASE COUNT         252 a        187 c        155 g        205 t          7 others
ORIGIN
```

```
Query Match            21.0%;  Score 79.4;  DB 12;  Length 806;
Best Local Similarity  98.8%;  Pred. No. 2.6e-12;
Matches   80;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;
```

Qy   262 GGTGATGGAGATGCCAATCCATGGAATCAGGTGTGGCCCAGCTATGTTGGTAGCTATAGCAG 321
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   423 GGTGATGGAGATGCCAATCCATGGAATCAGGTGTGGCCCACAGCTATGTTGGTAGCTATAGCAG 364
Qy   322 AAGTGTCTTCTTGGCAAAGATTC 342
           ||||||||||||||||||||||
Db   363 AAGTGTTCTTGGCAAAGATTC 343

```
RESULT 5
AQ035618
LOCUS          AQ035618               349 bp    DNA    linear    GSS 11-JUL-1998
DEFINITION     CIT-HSP-2319N3.TF CIT-HSP Homo sapiens genomic clone 2319N3, DNA
               sequence.
ACCESSION      AQ035618
VERSION        AQ035618.1      GI:3301715
KEYWORDS       GSS.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1  (bases 1 to 349)
AUTHORS        Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
               Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
               Simon,M. and Venter,J.C.
TITLE          Use of a random BAC End Sequence Database for Sequence-Ready Map
               Building (1998)
JOURNAL        Unpublished (1998)
COMMENT        Other_GSSs: CIT-HSP-2319N3.TR
               Contact: Mark Adams
               Department of Eukaryotic Genomics
               The Institute for Genomic Research
               9712 Medical Center Dr., Rockville, MD 20850, USA
               Tel: 301 838 0200
               Fax: 301 838 0208
               Email: mdadams@tigr.org
               Clones are available from Research Genetics (info@resgen.com). BAC
               end search page:
               http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
               Seq primer: M13-21
               Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..349
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2319N3"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT          85 a         69 c         84 g        111 t
ORIGIN
```

```
Query Match            16.6%;  Score 62.6;  DB 17;  Length 349;
Best Local Similarity  94.2%;  Pred. No. 1.7e-07;
Matches   65;  Conservative   0;  Mismatches   4;  Indels   0;  Gaps   0;
```

Qy   86 CAAGCTTGAATGTGTTCTGGAGCTGAAGGATGCACGGGTGTTAAGCCCGTGTTCTTTT 145
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31 CAAGCTTGAATGTGTGTTCTGGAGCTGAAGGATGCACGGGTGTTAAGCCCCGTGTTCTTTT 90
Qy   146 CCGTTGTTTT 154
           |||| ||
Db   91 CCGTAAGTT 99

```
RESULT 6
AQ035738/c
LOCUS          AQ035738               237 bp    DNA    linear    GSS 11-JUL-1998
DEFINITION     CIT-HSP-2321C21.TF CIT-HSP Homo sapiens genomic clone 2321C21, DNA
               sequence.
ACCESSION      AQ035738
VERSION        AQ035738.1      GI:3301835
KEYWORDS       GSS.
```

SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 237)
AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
TITLE   Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
JOURNAL   Unpublished (1998)
COMMENT   Contact: Mark Adams
            Other_GSSs: CIT-HSP-2321C21.TR
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES   Location/Qualifiers
     source 1..237
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="2321C21"
            /clone_lib="CIT-HSP"
            /sex="Male"
            /cell_type="Sperm"
            /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
            HindIII"

BASE COUNT   59 a   47 c   47 g   84 t
ORIGIN

Query Match   11.0%;   Score 41.4;   DB 17;   Length 237;
Best Local Similarity 97.7%;   Pred. No. 0.22;
Matches 42;   Conservative 0;   Mismatches 1;   Indels 0;   Gaps 0;

Qy   335 AAAGATTCCTCCCGGGAAGGAAGGTACCATTGGAGAAACCATGC 377
          |||||||||||| |||||||||| |||||||||||||||||||||
Db   237 AAAGATTCCTCCAGGGAAGGAAGGTACCATTGGAGAACCATGC 195

RESULT 7
LOCUS   BQ655384   554 bp   mRNA   linear   EST 15-JUL-2002
DEFINITION NXRV093_F11_F NXRV (Nsf Xylem Root wood Vertical) Pinus taeda cDNA
            clone NXRV093_F11 5', mRNA sequence.
ACCESSION   BQ655384
VERSION   BQ655384.1 GI:21787710
KEYWORDS   EST.
SOURCE   Pinus taeda
ORGANISM   loblolly pine
            Pinus taeda
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
REFERENCE   1 (bases 1 to 554)
AUTHORS   Sederoff,R.
TITLE   Molecular Basis of Wood Formation in the Pine Megagenome
JOURNAL   Unpublished (2000)
COMMENT   Contact: Johnson, Arthur
            North Carolina State University
            Tel: 919 515 7800
            Fax: 919 515 7801
            Email: ajohnson@unity.ncsu.edu
            Seq primer: T3.
FEATURES   Location/Qualifiers
     source 1..554
            /organism="Pinus taeda"
            /strain="Coastal plain loblolly pine from North Carolina"
            /db_xref="taxon:3352"
            /clone="NXRV093_F11"
            /clone_lib="NXRV (Nsf Xylem Root wood Vertical)"
            /tissue_type="Xylem"
            /cell_type="Root (primary)"
            /dev_stage="Transitional"
            /lab_host="XL1-Blue"
            /note="Vector: pBluescript SK-; Site_1: Eco RI; Site_2:
            XhoI; The library is from primary xylem scraped from the
            roots of a twelve year old tree in the transitional phase
            from juvenile wood to mature wood production. NOTE: The
            sequences contain a 'cDNA adapter' between the EcoRI site
            and the start of the EST. The adaptersequence is
            'AATTCGGCACGAG'."

BASE COUNT   153 a   88 c   131 g   161 t   21 others
ORIGIN

RESULT 8
LOCUS   BH853295   439 bp   DNA   linear   GSS 13-JUN-2002
DEFINITION BH853295 SALK_076311.54.00.x Arabidopsis thaliana TDNA Insertion lines
            Arabidopsis thaliana genomic clone SALK_076311.54.00.x, DNA
            sequence.
ACCESSION   BH853295
VERSION   BH853295.1 GI:21424166
KEYWORDS   GSS.
SOURCE   Arabidopsis thaliana
ORGANISM   Arabidopsis thaliana
            thale cress.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 439)
AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
TITLE   A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
JOURNAL   Unpublished (2001)
COMMENT   Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: eckergsalk.edu
            This is single pass sequence recovered from the left border of
            TDNA. This sequence lies within 300 bases of the 3' end of
            At2941810.
            Class: TDNA tagged.
FEATURES   Location/Qualifiers
     source 1..439
            /organism="Arabidopsis thaliana"
            /strain="Columbia 0"
            /db_xref="taxon:3702"
            /clone="SALK_076311.54.00.x"
            /clone_lib="Arabidopsis thaliana TDNA insertion lines"
            /note="PCR was performed on Arabidopsis thaliana TDNA insertion lines

Query Match   10.5%;   Score 39.8;   DB 14;   Length 554;
Best Local Similarity 53.5%;   Pred. No. 0.77;
Matches 83;   Conservative 0;   Mismatches 72;   Indels 0;   Gaps 0;

Qy   168 TTGGAATAAAAAACCTCCCGTGCCAAGTAGTACTTGGTTTTATGCTCAACACATGCTTTGACTG 227
          || ||| ||||| ||| ||||||||||| |||||||||||||||||||||| ||| ||||| ||
Db   287 TTCGAAAAGAAAAACATTACTGCCAAATTTTCCTGGTTTTGATGTTCGTTATTGGTTGGAAATG 346

Qy   228 TTGAAAAGAGAGACCTTTGGCCACACATTGAAGGGATGGTTGATGGAGATGCCAAATCCATGGA 287
          || |||| || | |||| || ||| || ||| || ||||| | || || ||||| | || |
Db   347 CAGAATAAGTTCTTCTTCCAGAAATACATGGGGTATGGGCACGGGTAAGCCAAATTATTGGA   406

Qy   288 TCAGGTGCGCGCAGCTATGTTGGTPAGCTATPAGCAGA 322
          | | ||| | || || |||| | || || |||| |
Db   407 TGAATCTGGGACGTTTGCCATGCTTCATTTGATPAGA   441

each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

```
BASE COUNT    155 a    68 c    68 g    146 t    2 others
ORIGIN

Query Match    10.4%;    Score 39.2;  DB 17;   Length 439;
Best Local Similarity  53.2%;  Pred. No. 1.1;
Matches  83;  Conservative  0;  Mismatches  73;  Indels  0;  Gaps  0;

Qy    106  GGAGCTGAAGGAGTGCCACGGTGTTGTGTAAGCCCCTGTTCTTTTCCGTTGTTGTTTAATCTAATTGTT 165
           |||||                              |||  || |||  | || || | ||
Db    123  GGAGACCATATATGGGACTTCTGGAACTCCCTGCCAAGTAGTAGTACTTGGTTTTTTTATGGTCTCACAAC 182

Qy    166  CTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTTTTATGGTCTCACACATGCTTTGAC 225
           |||| |        |||||||||||||  || ||||||||| || | || ||||| ||||||||||
Db    183  CATTTAATTAATTATAAATCTTAATTGACACAAAATTTCTTTGTTTTAACGTTTTTTTTTCTTTGTA 242

Qy    226  TGTTGAAAAGAGAGACCTTTGGCACACATTGAAGGGAT 261
           |||| ||  || |||||  ||||| ||| |||||||||
Db    243  AGTTAAACGTAAAGCTTTATTTAATAAACAAAAGAAACAAAT 278
```

RESULT 9

```
LOCUS        BI454874                1237 bp    mRNA    linear     EST 21-AUG-2001
DEFINITION   603173414F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5252883 5',
             mRNA sequence.
ACCESSION    BI454874
VERSION      BI454874.1  GI:15245530
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 1237)
AUTHORS      NIH-MGC http://mgc.ncl.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM1638  row: l  column: 04
             High quality sequence stop: 102.
FEATURES     Location/Qualifiers
   source    1..1237
             /organism="Mus musculus"
             /strain="C57BL/6J"
             /db_xref="taxon:10090"
             /clone="IMAGE:5252883"
             /clone_lib="NCI_CGAP_Mam5"
             /tissue_type="tumor, gross tissue"
             /dev_stage="7 months"
             /lab_host="DH10B"
             /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Sall;
             Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
             Library constructed by Life Technologies. Investigators
             providing samples: Lothar Hennighausen/Robin Humphreys,
             NIH"

BASE COUNT    431 a    226 c    336 g    243 t    1 others
ORIGIN

Query Match    10.3%;    Score 38.8;  DB 13;   Length 1237;
Best Local Similarity  50.3%;  Pred. No. 1.8;
Matches  94;  Conservative  0;  Mismatches  93;  Indels  0;  Gaps  0;
```

```
Qy     65  CACCCAAGCGCGGGAGCTCTCACAAGCTTGAAATGTGTGTTCTGGAGCTGAAGGAGTGCACGG 124
           ||            |||  ||||||| |         | |  | |||||||||||||||||||
Db    480  CCCCCCCCCTTTTCTTTTGTTGTTTGGTTGTTTGTTGTTGTATGTTGTTCTGTTTTGTTCTTTCTGTATC 421

Qy    125  TTGTTAACCCCCTGTTCTTTTTCCGTTGTTGTTTAATCTAATTGTTTTTTTTGGAATAAAAAACCTCC 184
           |||  | | |  |||||||    | |||||| ||||| |  | |  |  ||||| ||| ||  |||||
Db    420  TTCTCTATCTGTTTATTGTGTTTTCTTTTTGTTTATCGTGTTGTTTTTTTTTATATAPAGACA 361

Qy    185  CTGCCAAGTAGTACTTGGTTTTTTTATGGTCTCAACATGCTTTGACGTGTTGAAAAGAGAGACCTTTG 244
           |||||| |  | ||  ||||||||| | | || |||| || |||| |  | ||||| | | ||
Db    360  CAGCCAAPATATTTTTTTTAFTTGAGACACACCACTGTNTTTTGCCACATTAPACACACATTGG 301

Qy    245  GCACACA 251
           ||||||
Db    300  GCGCACA 294
```

RESULT 10

```
LOCUS        BQ392762                585 bp    mRNA    linear     EST 22-MAY-2002
DEFINITION   NISC_mg26004.y1 NICHD XGC Emb5 Silurana tropicalis cDNA clone
             IMAGE:5384863 5', mRNA sequence.
ACCESSION    BQ392762
VERSION      BQ392762.1  GI:21080449
KEYWORDS     EST.
SOURCE       western clawed frog.
ORGANISM     Silurana tropicalis
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
             Xenopodinae; Silurana.
REFERENCE    1  (bases 1 to 585)
AUTHORS      NIH-XCG http://image.llnl.gov/image/html/xenopuslib_info.shtml.
TITLE        National Institute of Child Health and Human Development, National
             Cancer Institute, Xenopus Gene Collection
JOURNAL      Unpublished (2002)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             cDNA Library Preparation:
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
             DNA Sequencing by: National Institutes of Health Intramural
             Sequencing Center (NISC)
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             info@image.llnl.gov
             Plate: LLAM11980  row: G   column: 8
             Seq primer: M13RP1 reverse primer (ABI).
FEATURES     Location/Qualifiers
   source    1..585
             /organism="Silurana tropicalis"
             /db_xref="taxon:8364"
             /clone="IMAGE:5384863"
             /clone_lib="NICHD XGC Emb5"
             /tissue_type="gastrula"
             /dev_stage="embryo, stages 10-13"
             /lab_host="DH10B (phage-resistant)"
             /note="Vector: pCMV-SPORT6.ccdb; Site_1: NotI; Site_2:
             EcoRV; Cloned unidirectionally. Primer: Oligo dT. Average
             insert size 2.0 kb. Constructed by Invitrogen. Note: This
             is a Xenopus Gene Collection (XGC) library."

BASE COUNT    141 a    109 c    163 g    172 t
ORIGIN

Query Match    10.1%;    Score 38;  DB 14;   Length 585;
Best Local Similarity  52.5%;  Pred. No. 2.6;
Matches  83;  Conservative  0;  Mismatches  75;  Indels  0;  Gaps  0;
```

```
Qy    144  TTCCGTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTACTTGGT 203
           ||||              |    ||    ||   ||||     || | |  | || |||||||||
Db    405  TTGCCATGGAGACTCTGATTTCCTTGGGATTTGAAAAGAGACACTTTGGCACACATTAGTGGCTGTGTGACA 464

Qy    204  TTTTATGCTTCAACATGTCTTTGACTGTTGAAAAGAGACACTCCTTGGCACACATTGAAGGGATGG 263
           ||||  |    |     | || |  ||  ||   | |  ||  ||||||||||||||||
```

RESULT 11

```
Db   465  CGTCTGCTTTGGAAGGTTTGCCTTTAATAAAGAGACTTGGTGAACAAGCCAAAGGACGAA  524
                                                                              
Qy   264  TGATGGAGGATGCCAATCCATGGAAATCAGGTGGCGCAGC  301
          |||||| ||||||||||||||||||||||||||||| |||
Db   525  TCATAGTAATGCCAGGTGGCGGCGCATTACTGACGGAAAC  562
```

LOCUS   AL628007   641 bp   mRNA   linear   EST 02-NOV-2001
DEFINITION   AL628007 XGC-gastrula Silurana tropicalis cDNA clone Tgas022e21 5',
                mRNA sequence.
ACCESSION   AL628007
VERSION   AL628007.1   GI:16597490
KEYWORDS   EST.
SOURCE   western clawed frog.
  ORGANISM   Silurana tropicalis
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
                Xenopodinae; Silurana.
REFERENCE   1 (bases 1 to 641)
  AUTHORS   Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
  TITLE   Sanger Xenopus tropicalis EST project 2001
  JOURNAL   Unpublished (2001)
COMMENT   Contact: Huckle E
                Sanger Centre
                Hinxton, Cambridgeshire, CB10 1SA, UK
                Email: trop@sanger.ac.uk
                Sanger Xenopus tropicalis EST project 2001
                TROPICALIS_SEQUENCE_ID: Tgas022e21.sp6
                Sequencing primer: SP6
                This sequence is from a Xenopus Gene Collection (XGC) library
                constructed by Aaron M. Zorn.
FEATURES   Location/Qualifiers
     source   1..641
                     /organism="Silurana tropicalis"
                     /db_xref="taxon:8364"
                     /clone="Tgas022e21"
                     /clone_lib="XGC-gastrula"
                     /dev_stage="gastrula (stages 10.5-13 mixed)"
                     /lab_host="Escherichia coli Xl1-blue"
                     /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
                     was oligo dT primed from 5ug of poly A+ RNA from stages
                     10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
                     into pCS107 with EcoRI at the 5' end and NotI at the 3'
                     end."
BASE COUNT   154 a   118 c   189 g   180 t
ORIGIN

RESULT 12

```
Qy   144  TTCCGTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTACTTGGT  203
          ||||||||| |||||||||||||||||||||||||   || ||||||||||||| |||||
Db   449  TTGCCATGGAGACTCTGATTTCCTTCGGATTGAAAGGGTTCTTACTTACTGGCTGTGGACA  508

Qy   204  TTTATGCTCAACATGCTTGACTGTGTGAAAAGAGAGACCTTTGCACACATTGAAGGGATGG  263
          ||||| ||||||||||||| |||||||||||| |||||||||||| ||||||||||||||
Db   509  CGTCTGCTTTGGAAGGTTTGCCTTTAATAAAGAGACTTGGTGAACAAGCCAAAGGACGAA  568

Qy   264  TGATGGAGGATGCCAATCCATGGAAATCAGGTGGCGCAGC  301
          ||||| |||||||||||||||||||||||||||| |||||
Db   569  TCATAGTAATGCCAGGTGGCGGCGCATTACTGACGGAAAC  606
```

LOCUS   AQ922388   592 bp   DNA   linear   GSS 21-DEC-1999
DEFINITION   RPCI-23-276B5.TJ RPCI-23 Mus musculus genomic clone RPCI-23-276B5,
                DNA sequence.
ACCESSION   AQ922388
VERSION   AQ922388.1   GI:6611391

Query Match   10.1%; Score 38; DB 9; Length 641;
Best Local Similarity   52.5%; Pred. No. 2.7;
Matches 83; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

KEYWORDS   GSS.
SOURCE   house mouse.
  ORGANISM   Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 592)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
                ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
                and Fraser,C.M.
  TITLE   Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT   Department of Eukaryotic Genomics
                The Institute for Genomic Research
                9712 Medical Center Dr., Rockville, MD 20850, USA
                Tel: 301 838 0200
                Fax: 301 838 0208
                Email: szhao@tigr.org
                Clones are derived from the mouse BAC library RPCI-23. For BAC
                library availability, please contact Pieter de Jong
                (pieter@dejong.med.buffalo.edu). Clones may be purchased from
                BACPAC Resources (http://bacpac.med.buffalo.edu/ordering/frame.htm)
                or from Resea ch Genetics (info@resgen.com). BAC end page:
                http://www.tigr.org/tdb/bac_ends/mouse/bac_ends_intro.html
                Plate: 276 row: B column: 5
                Seq primer: SP6
                Class: BAC ends.
FEATURES   Location/Qualifiers
     source   1..592
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-276B5"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and partially digested
                     with a combination of EcoRI and EcoRI Methylase. Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI sites. The ligation products were transformed into
                     DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT   161 a   115 c   130 g   186 t
ORIGIN

RESULT 13

```
Qy   109  GCTGAAGGAGTGCACGGTTGTTAAGCCCCTTGTTCTTTCCGTTGTTTTAACTCTAAGTTCTT  168
          ||       ||  ||||||||||||||         |||||||  |||    ||| |||
Db   398  GTTGAAATTTGCTCTTGCACTAGCATCCTGTTCTTTCTTCACCAGTTTTATCCACTTGTTTG  457

Qy   169  TGGAATAAAAAACCTCCCCTGCCAAGTAGTACTTGGTGTTTATGCTCAACATGCTTGACTGT  228
          ||||||| |||||  ||       ||||||||||||                 || |||
Db   458  TTAAATAACATGGCCAGTCTTAATTTGGAGACCCTTCTTTGGGCAGAATTTATTTGTAGT  517

Qy   229  TGAAAAGAGAGACCTTTGGCACACATTGAAGGGATGGTGATGGAGGATGCCAATCCATGGAAT  288
          |||     |||     |||||||||   |||   |||||| ||  || |||||| |||
Db   518  GGCCCAGGCCGCGGGTGACAGTAAGTGGGTCCTTGGGAGTGGTAGTGGTATTCTATTGTGTCT  577

Qy   289  CAG  291
          |||
Db   578  CAG  580
```

LOCUS   AZ108808   699 bp   DNA   linear   GSS 09-MAY-2000
DEFINITION   RPCI-23-25N18.TV RPCI-23 Mus musculus genomic clone RPCI-23-25N18,
                DNA sequence.
ACCESSION   AZ108808

Query Match   9.9%; Score 37.4; DB 17; Length 592;
Best Local Similarity   50.3%; Pred. No. 3.9;
Matches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

VERSION   AZ108808.1   GI:7761866
KEYWORDS   GSS.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 699)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levin,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE   Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT   Other_GSSs: RPCI-23-25N18.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Seq primer: T7
            Plate: 25 row: N column: 18
            Class: BAC ends.
FEATURES   Location/Qualifiers
     source 1..699
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="RPCI-23-25N18"
            /clone_lib="RPCI-23"
            /sex="Female"
            /lab_host="DH10B"
            /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
            EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
            brain genomic DNA was isolated and partially digested
            with a combination of EcoRI and EcoRI Methylase. Size
            selected DNA was cloned into the pBACe3.6 vector at the
            EcoRI sites. The ligation products were transformed into
            DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT   241 a   114 c   129 g   215 t
ORIGIN

Query Match   9.8%; Score 37; DB 17; Length 699;
Best Local Similarity   56.0%;   Pred. No. 5.3; Mismatches 55; Indels 0; Gaps 0;
Matches 70; Conservative 0;

Qy   121 ACGGTTGTTAAGCCCCCGTTCTTTTCCGTTGTTTAATCTAATGTTCTTTGGAATAAAAAC 180
             ||||||| ||||||||| |||||| ||| ||| | |||| |||||| ||||||
Db   538 ACTGCTGTGTTTCCTGCTTTCCTTCGAGCTTCCTCGATAAAAGCTAAAATTTATTTAGGA 479

Qy   181 CTCCCTGCCAAGTAGTACTTCGGTTTTTATGCTCCAACATGCTTTGACTGTTGAAAAGAGACC 240
             || | ||| ||| | | ||| ||| |||| ||| ||||| ||| ||
Db   478 CTTAGTCACATTGAAGACTTTATTCTGTCTCAACAAAATTTCTCTTATTTTTTCTAAAGACCCT 419

Qy   241 TTTTGG 245
             |||||
Db   418 TTTTGG 414

RESULT 14
LOCUS   AZ191123
DEFINITION   SP_1018_B2_D06_T7A Strongylocentrotus purpuratus, purple sea urchin
             , sperm genomic BAC library Strongylocentrotus purpuratus genomic
             clone Plate-1018 Col-12 Row-H, DNA sequence.
ACCESSION   AZ191123
VERSION   AZ191123.1   GI:8374302
LOCUS   AZ191123/c   829 bp   DNA   linear   GSS 30-AUG-2000

KEYWORDS   GSS.
SOURCE   Strongylocentrotus purpuratus.
  ORGANISM Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1 (bases 1 to 829)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and
            Hood,L.
  TITLE   A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT   Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 1018 row: H column: 12
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 829.
FEATURES   Location/Qualifiers
     source 1..829
            /organism="Strongylocentrotus purpuratus"
            /db_xref="taxon:7668"
            /clone="Plate-1018 Col-12 Row-H"
            /clone_lib="Strongylocentrotus purpuratus, purple sea
            urchin, sperm genomic BAC library"
            /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-coli
            DH10B"
BASE COUNT   258 a   190 c   151 g   229 t   1 others
ORIGIN

Query Match   9.8%; Score 37; DB 17; Length 829;
Best Local Similarity   54.5%;   Pred. No. 5.5; Mismatches 61; Indels 0; Gaps 0;
Matches 73; Conservative 0;

Qy   151 GTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTTATGC 210
             |||| | ||| | ||||| ||||||| || | | | |||||| ||| || ||||
Db   641 GTTTGATGTAACTNTTTTGTTGACTTGAAAAGAACCTTGGAGGTTATGCTTGAATTATCC 582

Qy   211 TCAACATGCTTTGACTGTTGAAAAAGAGACACCTTTGGCACACANTTGAAGGAGGATGGTGATGGA 270
             ||| | | || | ||| ||| || | ||||| || |||| | | ||| | |||
Db   581 GCCACACGTATTCAATATATGTTGAACAACTTTCTCGGGAACACACTTCAGGGAGGGGAGGGA 522

Qy   271 GATGCCAATCCATG 284
Db   521 GCTAGCCAACGCTG 508

RESULT 15
LOCUS   AI449346
DEFINITION   mt07d11.x1 Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:62373
             3', mRNA sequence.
ACCESSION   AI449346
VERSION   AI449346.1   GI:4289359
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 473)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.

LOCUS   AI449346   473 bp   mRNA   linear   EST 09-MAR-1999

```
TITLE       The WashU-NCI Mouse EST Project 1999
JOURNAL     Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            High quality sequence stop: 402.
FEATURES             Location/Qualifiers
     source          1..473
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:620373"
                     /clone_lib="Soares mouse 3NbMS"
                     /sex="male"
                     /tissue_type="spleen"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGGCGGCCGCCCTGTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through
                     three rounds of normalization, and was constructed by
                     Bento Soares and M.Fatima Bonaldo."
BASE COUNT      143 a      78 c      83 g     169 t
ORIGIN
```

```
Query Match          9.7%;    Score 36.8;    DB 9;    Length 473;
Best Local Similarity   54.4%;
Matches   74;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy    121 ACGGTTGTTAAGCCCCTGTTCTTTTCCGTTGTTAATCTAAATGTTCTTTGGAATAAAAAC 180
          ||  || |   |||| | || |      | | |    |   |||||| |||| |||| |
Db     48 ACTGTGCAATATTTCTTTTCACTATCCATGTCTATTTTTTATTTACTTGCTTGGAATAATATC 107

Qy    181 CTCCCTGCCAAGTAGTACTTGGTTTTTATGCTCACATGCTTTGACTGTTGAAAAGAGACC 240
          |||||| |  | ||| ||| |  | |    | |  |    ||| ||||   ||  ||||
Db    108 CTCACAGTAAAACCTTTTGTGGTTTTAATGACTTGCCATATGTTGCCAAATTAAAAGGATTCA 167

Qy    241 TTTTGGCACACATTGAA 256
          || ||| ||||||||||
Db    168 TTGGGCCTACACATTGAA 183
```

Search completed: January 31, 2003, 03:44:26
Job time : 2433 secs

OM nucleic - nucleic search, using sw model

```
                        GenCore version 5.1.3
                        Copyright (c) 1993 - 2003 Compugen Ltd.

Run On:        January 30, 2003, 22:59:19 ; Search time 263 Seconds
                                             (without alignments)
                        3236.712 Million cell updates/sec

Title:         US-09-803-719-222
Perfect score: 378
Sequence:      1 tacggctgcttaagacgact............taccattggagaacatgcg 378

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      2185239 seqs, 1125993159 residues

Total number of hits satisfying chosen parameters:    4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     1: N_Geneseq_101002:*
               2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
               3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
               4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
               5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
               6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
               7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
               8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
               9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
              10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
              11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
              12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
              13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
              14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
              15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
              16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
              17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
              18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
              19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
              20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
              21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
              22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
              23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
              24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
                  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	378	100.0	378	22	AAS37164	Novel human diagno
2	299.6	79.3	327	22	AAS37218	Novel human diagno
3	279.6	74.0	322	22	AAS37237	Novel human diagno
4	275	72.8	318	22	AAS37242	Novel human diagno
5	33.8	8.9	5378	24	ABL33878	Human immune syste
6	33.6	8.9	492	22	ABA08523	Human secreted pro
7	33.4	8.8	158	21	AAC29807	Human secreted pro
8	33.4	8.8	240	20	AA224099	Rat NF-1 DNA fragm
9	33.4	8.8	240	20	AA224100	Rat NF-1 DNA fragm
c 10	33.4	8.8	25509	21	AAB81467	N. meningitidis pa
c 11	33.4	8.8	349980	21	AAF21607	Neisseria meningit
c 12	33	8.7	303	22	AAK57669	Human Immune/haema
c 13	32.8	8.7	240	20	AA224101	Rat NF-1 DNA fragm
c 14	32.8	8.7	240	20	AA224102	Rat NF-1 DNA fragm
c 15	32.8	8.7	240	20	AA224103	Rat NF-1 DNA fragm
c 16	32.8	8.7	240	20	AA224104	Rat NF-1 DNA fragm
c 17	32.4	8.6	413	22	AA224105	Rat NF-1 DNA fragm
c 18	32.4	8.6	240	20	AAK53947	Murine translation
c 19	32.2	8.5	1445	16	AAQ90508	Tomato S-adenosyl-
c 20	32	8.5	410	21	ABQ60002	Human secreted pro
c 21	32	8.5	490	24	AAC26229	Human colon cancer
c 22	32	8.5	1604	22	AAF84752	Nucleotide sequenc
c 23	31.8	8.4	314	22	AA224098	Rat NF-1 DNA fragm
c 24	31.8	8.4	240	22	AAK70860	Human Immune/haema
c 25	31.8	8.4	606	23	ABV06997	Human prostate exp
c 26	31.8	8.4	3738	23	ABL21954	Human prostate exp
c 27	31.8	8.4	13317	22	AAS46681	Drosophila melanog
c 28	31.8	8.4	13317	24	AAS61366	Tumour suppressor
c 29	31.6	8.4	1003	23	AAS79681	DNA encoding novel
c 30	31.6	8.4	2966	24	ABL89380	Babesia microti an
c 31	31.6	8.4	3011	24	ABL89389	Babesia microti an
c 32	31.6	8.4	3011	19	AAV22740	Babesia microti BM
c 33	31.6	8.4	3701	20	AAX30004	Babesia microti an
c 34	31.6	8.4	3701	21	AAC65087	Babesia microti an
c 35	31.6	8.4	3701	24	ABL89331	Babesia microti an
c 36	31.6	8.4	3701	24	AAT58840	B. microti BMNI-10
c 37	31.6	8.4	580073	18	ABA61524	Babesia microti an
c 38	31.4	8.4	568	22	AAK09824	Mycoplasma genital
c 39	31.4	8.3	568	22	AAK35718	Human foetal liver
c 40	31.4	8.3	568	24	AAI41433	Human brain expres
c 41	31.4	8.3	568	23	ABS09989	Human bone marrow
c 42	31.4	8.3	1319	23	AAS87926	Probe #10119 used
c 43	31.2	8.3	40324	21	ABO67149	Human genome-deriv
c 44	31	8.2	362	21	AA67588	DNA encoding novel
c 44	31	8.2	362	21	AA67588	Human angiogenesis
c 44	31	8.2	362	21	AA67588	Eucalyptus grandis
c 45	30.8	8.1	470	23	ABV09451	Human prostate exp

ALIGNMENTS

```
RESULT 1
AAS37164
ID   AAS37164 standard; cDNA; 378 BP.
XX
AC   AAS37164;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Novel human diagnostic and therapeutic gene #222.
XX
KW   Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
OS   Homo sapiens.
XX
PN   WO200166753-A2.
XX
PD   13-SEP-2001.
XX
PF   09-MAR-2001; 2001WO-US07787.
XX
PR   09-MAR-2000; 2000US-0188609.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI   Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI   Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI   Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
DR   WPI; 2001-530177/58.
```

RESULT 2

PT New polynucleotides and polypeptides, useful for diagnosis and
PT treatment of breast, lung and colon cancer -
XX
PS Claim 1; Page 652; 1193pp; English.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lanson G;
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
DR WPI; 2001-530177/58.
XX
PT New polynucleotides and polypeptides, useful for diagnosis and
PT treatment of breast, lung and colon cancer -
XX
CC The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.
XX
SQ Sequence 378 BP; 100 A; 76 C; 101 G; 101 T; 0 other;

ID AAS37218;
XX
AC AAS37218;
XX
DT 17-DEC-2001 (first entry)
XX
DE Novel human diagnostic and therapeutic gene #276.
XX
KW Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
OS Homo sapiens.
XX
PN WO20016753-A2.
XX
PF 09-MAR-2001; 2001WO-US07787.
PD 13-SEP-2001.
PR 09-MAR-2000; 2000US-0188609.
XX
PA (CHIR) CHIRON CORP.

Query Match   100.0%; Score 378; DB 22; Length 378;
Best Local Similarity 100.0%; Pred. No. 2.8e-112;
Matches 378; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 TACGGCTGCTTAAGACGACTTAAGGGGGGAATGACCGCAGCCGCCTCTTAGAGGAACATATGG 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 TACGGCTGCTTAAGACGACTTAAGGGGGGAATGACCGCAGCCGCCTCTTAGAGGAACATATGG 60

Qy   61 AAAACACCCAAGCCGGAGTCTCTCACAAGCTTGAATGTGTGTTCTCGGAGCTGAAGGAATGC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAAACACCCAAGCCGGAGTCTCTCACAAGCTTGAATGTGTGTTCTCGGAGCTGAAGGAATGC 120

Qy  121 ACGGTTGTTAAGCCCCTGTTCTTTTCCGTTGTTTAATCTAATGTTCTTTGGAATAAAAAAC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ACGGTTGTTAAGCCCCTGTTCTTTTCCGTTGTTTAATCTAATGTTCTTTGGAATAAAAAAC 180

Qy  181 CTCCCTGCCAAGTAGTACTTGGTTTTATGTCTCAACAACATGCTTTGACTGTTGAAAAGACC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTCCCTGCCAAGTAGTACTTGGTTTTATGTCTCAACAACATGCTTTGACTGTTGAAAAGACC 240

Qy  241 TTTGGCACACATTGAAGGGATGGTGATGGTGAGATGCCAATCCAATGGAATCAGGTGGCGCAG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TTTGGCACACATTGAAGGGATGGTGATGGTGAGATGCCAATCCAATGGAATCAGGTGGCGCAG 300

Qy  301 CTATGTTGGTGAGCTATAGCAGGAGAGTCTTCTTGGCAAAGAATTCCTCCCGGGAAGGAAGGTA 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CTATGTTGGTGAGCTATAGCAGGAGAGTCTTCTTGGCAAAGAATTCCTCCCGGGAAGGAAGGTA 360

Qy  361 CCATTGGAGAACCATCGG 378
        ||||||||||||||||||
Db  361 CCATTGGAGAACCATCGG 378
```

RESULT 3

ID AAS37237;
XX
AC AAS37237;
XX
DT 17-DEC-2001 (first entry)
XX
DE Novel human diagnostic and therapeutic gene #295.
XX
KW Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
OS Homo sapiens.
XX
PN WO20016753-A2.
XX
PD 13-SEP-2001.

Query Match   79.3%; Score 299.6; DB 22; Length 327;
Best Local Similarity 98.7%; Pred. No. 6.8e-87;
Matches 302; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy   28 GAAGTGACGCAGCGCGGCCTCTTAGAGGAACATATGGAAAAACACCCAAGCCGGAGTCTCTCACA 87
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 GAATGACGCAGCAGCGTCTTGAGAGGAACATAAGGAAAAACACCCAAGCCGGAGTCTCTCACA 80

Qy   88 AGCTTGAATGTGTGTTCTGGAGCTGAAGGAATGCACGGTTGTTAAGCCCCTGTTTCTTTTCC 147
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 AGCTTGAATGTGTGTTCTGGAGCTGAAGGAATGCACGGTTGTTAAGCCCCTGTTTCTTTTCC 140

Qy  148 GTTGTTTAATCTAATGTTCTTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA 207
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  141 GTTGTTTAATCTAATGTTCTTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA 200

Qy  208 TGCTCAACAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 267
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  201 TGCTCAACAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 260

Qy  268 GGAGATGCCAAATCCAATGGAATCAGGTGGCGCAGCTATGTTGGTGAGCTATAGCAGAAGTCT 327
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  261 GGAGATGCCAAATCCAATGGAATCAGGTGGCGCAGCTATGTTGGTGAGCTATAGCAGAAGTCT 320

Qy  328 TCTTGG 333
        ||||||
Db  321 TCTTGG 326
```

PF 09-MAR-2001; 2001WO-US07787.
XX
PR 09-MAR-2000; 2000US-0188609.
XX
PA (CHIR) CHIRON CORP.
    (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
DR WPI; 2001-530177/58.
XX
PT New polynucleotides and polypeptides, useful for diagnosis and
PT treatment of breast, lung and colon cancer -
XX
PS Claim 1; Page 670; 1193pp; English.
XX
CC The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.
XX
SQ Sequence 322 BP; 90 A; 64 C; 78 G; 89 T; 1 other;

Query Match   74.0%; Score 279.6; DB 22; Length 322;
Best Local Similarity   98.6%; Pred. No. 2e-80;
Matches 282; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy   28 GAATGACGCAGCGCGGCTCTTAGAGGAAGACATATGGAAAACACCCAAGCCGGAGTCTCTCACA 87
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36 GAATGACGCAGCAGCTCTGAGAGGAACATAAGGAAAAACACCCAAGCCGGAGTCTCTCACA 95

Qy   88 AGCTTGAATGTGTGTTCTGGAGCTGAAGGAATGCACGGTTGTTAAGCCCCTGTTCTTTTCC 147
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   96 AGCTTGAAATGTGTGTTCTGGAGCTGAAGGAATGCACGGTTGTTAAGCCCCTGTTCTTTTCC 155

Qy 148 GTTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCGCCAAGTACTTGGTTTTA 207
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 156 GTTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCGCCAAGTAGTACTTGGTTTTA 215

Qy 208 TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 267
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 216 TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 275

Qy 268 GGAGATGCCAATCCATGGAATCAGGTGGGCGCCAGCTATGTTGGTAGC 313
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db 276 GGAGATGCCAATCCATGGAATCAGGTGGGCGCCAGCTATGTTGGTAGC 321

RESULT 4
AAS37242
ID   AAS37242 standard; cDNA; 318 BP.
XX
AC   AAS37242;
XX
DT   17-DEC-2001 (first entry)
XX
DE   Novel human diagnostic and therapeutic gene #300.
XX
OS   Homo sapiens.
XX
KW   Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
PN   WO200166753-A2.
XX
PD   13-SEP-2001.

XX
PF 09-MAR-2001; 2001WO-US07787.
XX
PR 09-MAR-2000; 2000US-0188609.
XX
PA (CHIR) CHIRON CORP.
    (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
DR WPI; 2001-530177/58.
XX
PT New polynucleotides and polypeptides, useful for diagnosis and
PT treatment of breast, lung and colon cancer -
XX
PS Claim 1; Page 671; 1193pp; English.
XX
CC The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.
XX
SQ Sequence 318 BP; 90 A; 63 C; 76 G; 89 T; 0 other;

Query Match   72.8%; Score 275; DB 22; Length 318;
Best Local Similarity   98.2%; Pred. No. 6.3e-79;
Matches 278; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy   28 GAATGACGCAGCGCGGCTCTTAGAGGAAGACATATGGAAAACACCCAAGCCGGAGTCTCTCACA 87
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36 GAATGACGCAGCAGCTCTGAGAGGAACATAAGGAAAAACACCCAAGCCGGAGTCTCTCACA 95

Qy   88 AGCTTGAATGTGTGTTCTGGAGCTGAAGGAATGCACGGTTGTTAAGCCCCTGTTCTTTTCC 147
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   96 AGCTTGAAATGTGTGTTCTGGAGCTGAAGGAATGCACGGTTGTTAAGCCCCTGTTCTTTTCC 155

Qy 148 GTTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCGCCAAGTAGTACTTGGTTTTA 207
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 156 GTTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCGCCAAGTAGTACTTGGTTTTA 215

Qy 208 TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAAGGGATGGTGAT 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 216 TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 275

Qy 268 GGAGATGCCAATCCATGGAATCAGGTGGGCGCCAGCTATGTTGGT 310
         |||||||||||||||||||||||||||||||||||||||||||||
Db 276 GGAGATGCCAATCCATGGAATCAGGTGGGCGCCAGCTATGTTGGT 318

RESULT 5
ABL33878
ID   ABL33878 standard; DNA; 5378 BP.
XX
AC   ABL33878;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1851.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; antiasthmatic;
KW   neuroprotective; anti-HIV; anticonvulsant; nootropic;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO020200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation -
XX
PS   Claim 1; SEQ ID NO 1851; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 5378 BP; 1499 A; 134 C; 1201 G; 2544 T; 0 other;

Query Match   8.9%;   Score 33.8;   DB 24;   Length 5378;
Best Local Similarity   48.2%;   Pred. No. 3.1;
Matches   95;   Conservative   0;   Mismatches 102;   Indels   0;   Gaps   0;

Qy   137 TGTTCTTTTCCGTTGTTTAATCTAAGTGTTCTTTGGAATAAAAACCTCCCGCCAAGTAGT 196
          ||| ||| ||||||| || | ||| |||||| || || ||||||| || ||||| ||
Db 1377 TTTTAATTTTAGTGGTTAAATTTATTTTTATTTTGAAATTTTATTATTTTAATATTAGTGGT 1436

Qy   197 ACTTGGGTTTTATGCTCAACATGCTTTGACTGTTGAAAAAGAGACCTTTGGCACACATTGAA 256
          || ||| || ||| || ||||| ||||| ||| | ||| | |||
Db 1437 ACGAAGTTTAGTGATTTATTTCGGATCCGTTGTTGGAGAATGGAGGTTTTTTAAATTATTAAG 1496

Qy   257 GGGATGGTGATCGAGATGCCAATCCATGGAATCCAGGTGCGCGAGCTACGTTGGTAGCTTAT 316
          || ||| | ||| ||||||| | || || || | ||||| | | |||| ||
Db 1497 ATAATTTTATTTTAATTTATTTTTTAGGGGTGAAAGTTTTTTTTGTTTTTTAAGGGAT 1556

Qy   317 AGCAGAAGTCTTCTTGG 333
          ||| || | || |||
Db 1557 AATTGAGTAATATTAG 1573

RESULT 6
ABA08523
ID   ABA08523 standard; cDNA; 492 BP.
XX
AC   ABA08523;
XX
DT   11-JAN-2002 (first entry)
XX
DE   Human secreted protein homologue-encoding cDNA, SEQ ID NO:299.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW   haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW   inhibin; chemotaxis; chemokinesis; oncogenesis;
KW   proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW   myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW   chronic inflammatory condition; proliferative retinopathy;
KW   atherosclerosis; coronary heart disease; arterial ischaemia;
KW   bone disorder; osteoporosis; vascular growth disorder;
KW   tissue regeneration; wound healing; infection; immune disorder;
KW   cell culture; drug screening; gene therapy; antiinflammatory;
KW   antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW   cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW   antifungal; vulnerary; antiulcer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157188-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03800.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
PA   (HYSE-) HYSEQ INC.
XX
DR   WPI; 2001-457740/49.
DR   P-PSDB; ABB11279.
XX
PT   Human proteins and DNA encoding sequences useful for preventing,
PT   treating or ameliorating a medical condition in a mammalian subject
PT   e.g. arthritis and cancer -
XX
PS   Claim 1; Page 440; 1963pp; English.
XX
CC   Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC   sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC   invention also relates to vectors and recombinant host cells comprising a
CC   nucleotide of the invention, methods of producing the novel polypeptides,
CC   antibodies against the polypeptides, methods of detecting the nucleotides
CC   or polypeptides in a sample, and methods of identifying compounds which
CC   bind to polypeptides of the invention. Although novel, many of the
CC   polypeptides of the invention have homology to known proteins, thereby
CC   giving an insight into their probable biological activities, and hence
CC   potential therapeutic applications. The polypeptides of the invention may
CC   have various activities, including cytokine; cell proliferation or cell
CC   differentiation activities; stem cell growth factor activity;
CC   haematopoiesis regulatory activity; tissue growth activity;
CC   immunomodulatory activity; activin- or inhibin-related activities;
CC   chemotactic or chemokinetic activities; receptor or ligand activities;
CC   thrombolytic activities; haemostatic; thrombotic or
CC   involved in oncogenesis, cancer cell proliferation or metastasis.
CC   Depending on their biological activities, polypeptides and nucleotides of
CC   the invention are useful for preventing, treating or ameliorating medical
CC   conditions, e.g., by protein or gene therapy. Such conditions include
CC   cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
CC   disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC   proliferative retinopathy, atherosclerosis, coronary heart disease,
CC   arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC   vascular growth. Polypeptides involved with tissue regeneration and
CC   repair (or nucleic acids encoding them) may be used to promote wound
CC   healing (e.g., of burns, incisions and ulcers), while those with
CC   immunomodulatory activities may be used in the treatment of viral,
CC   bacterial and fungal infections in addition to immune disorders.
CC   Polypeptides with growth factor activity may be used in cell cultures to
CC   promote cell growth. For example, such polypeptides may be used to
CC   manipulate stem cells in culture to give rise to neurepithelial cells
CC   that can be used to augment or replace cells damaged by illness,
CC   autoimmune disease or accidental damage. The polypeptides and nucleotides
CC   may also be used in the diagnosis of the above conditions, and in drug
CC   screening techniques. The present sequence represents a cDNA encoding a
CC   novel human polypeptide of the invention.
XX
SQ   Sequence 492 BP; 124 A; 97 C; 99 G; 172 T; 0 other;

Query Match
Best Local Similarity   52.9%;   Score 33.6;   DB 22;   Length 492;
Matches   72;   Conservative   0;   Pred. No. 1.3;   Mismatches   64;   Indels   0;   Gaps   0;

```
Qy    91 TTGAATGTGTTGTTCTGGAGCTGAAGGAGATGCACGGTTGTTAAGCCCCTGTTCTTTTCCGTT 150
         | || |    |  |   |||      ||  |    |  ||| ||| |  |      |||| | |
Db    51 TGGAACCTCACGCTCTTGGTTTGGTTGGAGTCATCAGATACCTGGGTTGTTATTCCAGGC 110

Qy   151 GTTTAATCTAANTGTTCTTTGGAATAAAAAACCTCCGTCGCCAAGTAGTACTACGTTTTATG 210
         |||| |  | |  | | | | ||||   |  |    ||   ||    |   ||  |  | |||
Db   111 ATATAATGTGTCGATTTTTAACAATGCAGGCCTCACTGCCAAAAAGTTCTTCGGTTTTTGTGC 170

Qy   211 TCAACATGCTTTGACT 226
         | |||  |  | | ||
Db   171 TTGTGCTCGGTATGATT 186
```

RESULT 7
ID   AAC29807/c
AC   AAC29807 standard; cDNA; 158 BP.
XX
AC   AAC29807;
XX
DT   06-OCT-2000 (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 33882.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP103401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999; 99US-0122487.
XX
PA   (GENS) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 33882; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 158 BP; 35 A; 34 C; 34 G; 55 T; 0 other;

Query Match
Best Local Similarity   53.4%;   Score 33.4;   DB 21;   Length 158;
Matches   70;   Conservative   0;   Pred. No. 0.9;   Mismatches   61;   Indels   0;   Gaps   0;

```
Qy     1 TACGGCTCCTTAAGACGACTTAAGGGGGAATGACGCACGCGCTCTTAGAGGAACAATATGG 60
         | ||| |    || | | |  |||    |  ||| |  | | |   |||  |  | | |||
Db   144 TATAGCTCCTGAGGAGGCGCACAAAAGGGAAATCTCCATTCATCTCTTTCTTCGCTCACAGAGA 85

Qy    61 AAAACACCCAAGCGCGGAGTCTCTCACAAGCTTGAAATGTGTTGTTCTGGAGCTGAAGGATGC 120
         || || || ||||| |||   |  |  |||  | |  | | || ||  ||| ||  ||   ||
Db    84 ACAGCATTAATCCTAGAGCCCACTAACACTCTGGAGAGTGAAATGGTGGGCTGACTGAAGA 25

Qy   121 ACGGTTGTTAA 131
         |  | | | ||
Db    24 AGGGGAGAAAA 14
```

RESULT 8
ID   AAZ24099/c
AC   AAZ24099 standard; DNA; 240 BP.
XX
AC   AAZ24099;
XX
DT   04-FEB-2000 (first entry)
XX
DE   Rat NF-1 DNA fragment 2.
XX
KW   NF-1; rat; primer; mutagenicity; mutation rate; polymerase error rate;
KW   radiation; ss.
XX
OS   Rattus sp.
XX
PN   DE19818422-A1.
XX
PD   28-OCT-1999.
XX
PF   24-APR-1998; 98DE-1018422.
XX
PR   24-APR-1998; 98DE-1018422.
XX
PA   (HARJ/) HARJES S.
XX
PI   Harjes S;
XX
DR   WPI; 1999-602407/52.
XX
PT   Method for determining absolute mutagenicity and error rate of
PT   polymerases in amplification reactions -
XX
PS   Examples; Fig 5; 18pp; German.
XX
CC   This invention describes a novel method for determining the absolute
CC   mutagenicity of a substance. The method is used to determine the
CC   mutation rate associated with particular substances and the error rates
CC   for polymerases during amplification reactions. It is therefore useful
CC   for monitoring nutrients or medicines and environmental factors such as
CC   radiation from televisions or computer screens. The method is applicable
CC   to all types of mutations (collectively or individually), is unaffected
CC   by repair mechanisms or in vivo selection mechanisms, and is an in vitro
CC   process so animal testing is not required. AAZ24098-AAZ24105 represent
CC   rat NF-1 DNA fragments used to illustrate the method of the invention.
XX
SQ   Sequence 240 BP; 66 A; 53 C; 53 G; 68 T; 0 other;

Query Match
Best Local Similarity   72.9%;   Score 33.4;   DB 20;   Length 240;
Matches   43;   Conservative   0;   Pred. No. 1.1;   Mismatches   16;   Indels   0;   Gaps   0;

```
Qy   148 GTTGTTTAATCTAANTGTTCTTTGGAATAAAAAACCTCCGTCGCCAAGTAGTACTACGTTTT 206
         ||| |     |||| | | |   || |||||||||||||||||||||  ||| | ||| ||||
Db    60 GTGGGACAATCTGATGCTATATCTTAAGAAAAAACCTCCGTCGCAAGTTCGAAGTTGCTTTT 2
```

RESULT 9
ID   AAZ24100/c
AC   AAZ24100 standard; DNA; 240 BP.
XX

AC   AA224100;
XX
DT   04-FEB-2000 (first entry)
XX
DE   Rat NF-1 DNA fragment 3.
XX
KW   NF-1; rat; primer; mutagenicity; mutation rate; polymerase error rate;
KW   radiation; ss.
XX
OS   Rattus sp.
XX
PN   DE19818422-A1.
XX
PD   28-OCT-1999.
XX
PF   24-APR-1998; 98DE-1018422.
XX
PR   24-APR-1998; 98DE-1018422.
XX
PA   (HARJ/) HARJES S.
PI   Harjes S;
XX
DR   WPI; 1999-602407/52.
XX
PT   Method for determining absolute mutagenicity and error rate of
PT   polymerases in amplification reactions -
PS   Examples; Fig 5; 18pp; German.
XX
CC   This invention describes a novel method for determining the absolute
CC   mutagenicity of a substance. The method is used to determine the
CC   mutation rate associated with particular substances and the error rates
CC   for polymerases during amplification reactions. It is therefore useful
CC   for monitoring nutrients or medicines and environmental factors such as
CC   radiation from televisions or computer screens. The method is applicable
CC   to all types of mutations (collectively or individually), is unaffected
CC   by repair mechanisms or in vivo selection mechanisms, and is an in vitro
CC   process so animal testing is not required. AA224098-224105 represent
CC   rat NF-1 DNA fragments used to illustrate the method of the invention.
XX
SQ   Sequence 240 BP; 66 A; 53 C; 53 G; 68 T; 0 other;

Query Match   8.8%; Score 33.4; DB 20; Length 240;
Best Local Similarity 72.9%;   Pred. No. 1.1;
Matches   43; Conservative   0; Mismatches 16; Indels   0; Gaps   0;

Qy   148 GTTGTTTAATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTAGTTGGTTTTT 206
              ||| ||||| |||| |||||||||||||||||||||||||||||||| |||||||| |||
Db   60 GTGGGACAATCTCATGCTATATCTAAGAAAAAACCTCCGTGCCAAGTCGAAGTTGCTTTTT 2

RESULT 10
AAA81467/c
ID   AAA81467 standard; DNA; 25509 BP.
XX
AC   AAA81467;
XX
DT   04-DEC-2000 (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_15 SEQ ID NO:15.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999; 99WO-US23573.
XX
PR   09-OCT-1998; 98US-0103794.
PR   30-APR-1999; 99US-0132068.
XX
PA   (CHIR) CHIRON CORP.
XX
PI   Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI   Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI   Rappuoli R, Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Claim 7; Page 399-406; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neisseriae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 25509 BP; 6157 A; 6156 C; 6549 G; 6647 T; 0 other;

Query Match   8.8%; Score 33.4; DB 21; Length 25509;
Best Local Similarity 52.5%;   Pred. No. 8.4;
Matches   73; Conservative   0; Mismatches 66; Indels   0; Gaps   0;

Qy   169 TGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTTATGCTTCAACATGCTTTGACTGT 228
              || ||||||| |||||||||||||||||| || ||||| || | |||||| || |||
Db   23305 TGATATAAAAAATCATCCAAGAGAAAACATTAATAATTTTGGAATAATGTCCCTTTAATAT 23246

Qy   229 TGAAAAGAGACCTTTGGCACACATTGAAGGGATTGGTGATGGAGATGCCAATCCATGGAAT 288
              ||| || || ||| | | | || ||||| |||| ||||| | |||| | |||| || |||
Db   23245 CGAAGAGAAAAGAAATCTCCAAAAATATATTAAAAGCGGGTGGGGTTTTCCGACAAATAGTTT 23186

Qy   289 CAGGTGGCGCAGCCTATGTT 307
              ||| |||| ||| || ||
Db   23185 TAATTCGGAGCAGCTATATT 23167

RESULT 11
AAF21607/c
ID   AAF21607 standard; DNA; 349980 BP.
XX
AC   AAF21607;
XX
DT   13-MAR-2001 (first entry)
XX
DE   Neisseria meningitidis B nucleotide sequence SEQ ID NO:108.
XX

RESULT 12

```
Db    75748 TAATTGGAGCAGCTATATT 75730
         |||  ||| ||||||||| ||
Qy     289 CAGGTGCGCGCAGCTATGTT 307

Db    75808 CCAAGAGAAAGAATCTCCAAAAATATATATAAAAGCGGGTGGGTTTTCCCGACAAAATAGTTT 75749
         ||  ||| ||| ||  |  |||       |||       |||      ||| ||  |||  ||
Qy     229 TGAAAAGAGACCTTTGGCACACATTGAAGGATGGGTGATGGAGAGATGCCAATCCATGGAAT 288

Db    75868 TGATATAAAAAATCATCCAAGAAAAACATTAATAATTTTGGAATAATGTCCCTTTAATAT 75809
         ||  |||||| |||  |||||||  || |||||||||||  ||||| |||  |  ||| |
Qy     169 TGGAATAAAAACCTCCCTGCCAAGTACTACTCTGGTTTTATGCTCACATGCTTTGACTGT 228
```

Query Match   8.8%; Score 33.4; DB 21; Length 349980;
Best Local Similarity   52.5%; Pred. No. 27;
Matches 73; Conservative   0; Mismatches 66; Indels   0; Gaps   0;

SQ   Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;

XX
CC   used.
CC   more effective in vaccines than the outer membrane proteins currently
CC   further opportunities to find antigenic or immunogenic proteins which are
CC   or coding sequences within the NMB genome. The DNA sequences provide
CC   databases can be used in a search to identify open reading frames (ORFs)
CC   bacteria. Computers, computer storage medium or computer
CC   presence of Neisserial bacteria or of antibodies raised to Neisserial
CC   due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC   proteins can be used in compositions for treating or preventing infection
CC   Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC   antibacterial activity, and can be used in vaccines and gene therapy.
CC   the present invention. The NMB genome and fragments from it have
CC   AAF21606 represent PCR primers which are used in the exemplification of
CC   Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC   the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC   sequence by 49980 bp (i.e. the last 49980 bp of AAF21607 are repeated at
CC   the beginning of AAF21607, the last 49980 bp of AAF21544 is repeated at
CC   sequences which overlap each other at the beginning and end of each
CC   sequence was too long to go in a record on its own it was split into 8
CC   to AAF21613 represent fragments of the NMB genomic sequence, as the
CC   Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC   The present invention describes the full length genome of
XX
PS   Claim 7; Appendix A; 692pp; English.
XX
PT   frames are used to detect, treat and prevent Neisserial infections -
PT   Neisseria meningitidis B full length genome sequence and open reading
XX
DR   WPI; 2000-647603/62.
PI   Frazer CM, Grandi G;
PI   Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI   Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
XX
PA   (GENO-) INST GENOMIC RES.
PA   (CHIR) CHIRON CORP.
XX
PR   28-FEB-2000; 2000GB-0004695.
PR   08-OCT-1999; 99WO-US23573.
PF   08-OCT-1999; 99US-0132068.
XX
PN   08-MAR-2000; 2000WO-US05928.
XX
PD   09-NOV-2000.
XX
PN   WO200066791-A1.
XX
OS   Neisseria meningitidis.
XX
KW   ds.
KW   diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW   Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
```

```
AC    AAK57669;
XX
ID    AAK57669 standard; cDNA; 303 BP.
AAK57669/c
```

XX
DT   06-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:2729.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0215647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0227759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0229924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.

```
PR   14-SEP-2000;   2000US-0232398.
PR   14-SEP-2000;   2000US-0232399.
PR   14-SEP-2000;   2000US-0232400.
PR   14-SEP-2000;   2000US-0232401.
PR   14-SEP-2000;   2000US-0233063.
PR   14-SEP-2000;   2000US-0233064.
PR   14-SEP-2000;   2000US-0233065.
PR   14-SEP-2000;   2000US-0234223.
PR   21-SEP-2000;   2000US-0234274.
PR   21-SEP-2000;   2000US-0234997.
PR   25-SEP-2000;   2000US-0234998.
PR   26-SEP-2000;   2000US-0235484.
PR   27-SEP-2000;   2000US-0235834.
PR   27-SEP-2000;   2000US-0235836.
PR   29-SEP-2000;   2000US-0236327.
PR   29-SEP-2000;   2000US-0236367.
PR   29-SEP-2000;   2000US-0236368.
PR   29-SEP-2000;   2000US-0236369.
PR   29-SEP-2000;   2000US-0236370.
PR   02-OCT-2000;   2000US-0237037.
PR   02-OCT-2000;   2000US-0237038.
PR   02-OCT-2000;   2000US-0237039.
PR   02-OCT-2000;   2000US-0237040.
PR   13-OCT-2000;   2000US-0239935.
PR   13-OCT-2000;   2000US-0239937.
PR   20-OCT-2000;   2000US-0240960.
PR   20-OCT-2000;   2000US-0241221.
PR   20-OCT-2000;   2000US-0241785.
PR   20-OCT-2000;   2000US-0241786.
PR   20-OCT-2000;   2000US-0241787.
PR   20-OCT-2000;   2000US-0241808.
PR   20-OCT-2000;   2000US-0241809.
PR   20-OCT-2000;   2000US-0241826.
PR   01-NOV-2000;   2000US-0244617.
PR   01-NOV-2000;   2000US-0244621.
PR   08-NOV-2000;   2000US-0246474.
PR   08-NOV-2000;   2000US-0246475.
PR   08-NOV-2000;   2000US-0246476.
PR   08-NOV-2000;   2000US-0246477.
PR   08-NOV-2000;   2000US-0246478.
PR   08-NOV-2000;   2000US-0246523.
PR   08-NOV-2000;   2000US-0246524.
PR   08-NOV-2000;   2000US-0246525.
PR   08-NOV-2000;   2000US-0246526.
PR   08-NOV-2000;   2000US-0246527.
PR   08-NOV-2000;   2000US-0246528.
PR   08-NOV-2000;   2000US-0246532.
PR   08-NOV-2000;   2000US-0246609.
PR   08-NOV-2000;   2000US-0246610.
PR   08-NOV-2000;   2000US-0246611.
PR   08-NOV-2000;   2000US-0246613.
PR   08-NOV-2000;   2000US-0249207.
PR   08-NOV-2000;   2000US-0249208.
PR   17-NOV-2000;   2000US-0249209.
PR   17-NOV-2000;   2000US-0249210.
PR   17-NOV-2000;   2000US-0249211.
PR   17-NOV-2000;   2000US-0249212.
PR   17-NOV-2000;   2000US-0249213.
PR   17-NOV-2000;   2000US-0249214.
PR   17-NOV-2000;   2000US-0249215.
PR   17-NOV-2000;   2000US-0249216.
PR   17-NOV-2000;   2000US-0249217.
PR   17-NOV-2000;   2000US-0249218.
PR   17-NOV-2000;   2000US-0249244.
PR   17-NOV-2000;   2000US-0249245.
PR   17-NOV-2000;   2000US-0249264.
PR   17-NOV-2000;   2000US-0249265.
PR   17-NOV-2000;   2000US-0249297.
PR   17-NOV-2000;   2000US-0249299.
PR   17-NOV-2000;   2000US-0249300.
PR   17-NOV-2000;   2000US-0249300.
PR   01-DEC-2000;   2000US-0250160.
PR   01-DEC-2000;   2000US-0250391.
PR   05-DEC-2000;   2000US-0251030.

PR   05-DEC-2000;   2000US-0251988.
PR   05-DEC-2000;   2000US-0256719.
PR   06-DEC-2000;   2000US-0251479.
PR   08-DEC-2000;   2000US-0251856.
PR   08-DEC-2000;   2000US-0251868.
PR   08-DEC-2000;   2000US-0251869.
PR   08-DEC-2000;   2000US-0251989.
PR   08-DEC-2000;   2000US-0251990.
PR   11-DEC-2000;   2000US-0254097.
PR   05-JAN-2001;   2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-483426/52.
DR   P-PSDB; AAM84888.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Claim 1; SEQ ID NO 2729; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 303 BP; 100 A; 62 C; 45 G; 94 T; 2 other;

Query Match             8.7%;  Score 33;  DB 22;  Length 303;
Best Local Similarity   48.1%;
Matches   90;  Conservative   1;  Mismatches   96;  Indels   0;  Gaps   0;
                                                Pred. No 1.6;

Qy   20 TTAAGGGGGAATGACGCAGCGGCGCTCTTAGAGGAAACATATGGAAAACACCCAAGCCGGAGT 79
            | ||| || || |  ||  | |   | | | | ||| | | ||||| | | |||
Db  297 TTGGGGTGGGGAGGCTGCCCCAGTGCAGAAATGCCAATTAAGAAAAACAAANTGTACAGAACC 238

Qy   80 CTCTCACAAGCTTGAATGTGTGGTTCTCGGAGCTGAAGGATGCACGGGTTGTTAAGCCCCTGT 139
           |||| |||  || | |  |  | ||| | || | || | | |||| | || ||||| |
Db  237 CTCTCTCACACATGAATTCATGCTTGTCTCGCCCAATCCTGTGAAAGTATTATTATAGGTCTCCT 178

Qy  140 TCTTTTCCGTTGTTGTTTTAATCTAATGTTCTTTTGGAATAAAAACCTCCCTGCCAAGTAGTGACT 199
          | ||| |  | ||  ||| | | |||| | || |||| ||||| | ||||||| |  ||| ||
Db  177 TGTTTCTCAGGGGTTCGGTGAATGCTCTCTTTTAAGGTTAACATGAATAGACAAWAGTTCT 118

Qy  200 TGGTTTT 206
           |||
Db  117 GGTCTTT 111

RESULT 13
AAZ24101/c
ID   AAZ24101 standard; DNA; 240 BP.
XX
AC   AAZ24101;
XX
DT   04-FEB-2000 (first entry)
XX
```

```
DE   Rat NF-1 DNA fragment 4.
XX
KW   NF-1; rat; primer; mutagenicity; mutation rate; polymerase error rate;
KW   radiation; ss.
XX
OS   Rattus sp.
XX
PN   DE19818422-A1.
XX
PD   28-OCT-1999.
XX
PF   24-APR-1998;  98DE-1018422.
XX
PR   24-APR-1998;  98DE-1018422.
XX
PA   (HARJ/) HARJES S.
XX
PI   Harjes S;
XX
DR   WPI; 1999-602407/52.
XX
PT   Method for determining absolute mutagenicity and error rate of
PT   polymerases in amplification reactions -
XX
PS   Examples; Fig 5; 18pp; German.
XX
CC   This invention describes a novel method for determining the absolute
CC   mutagenicity of a substance. The method is used to determine the
CC   mutation rate associated with particular substances and the error rates
CC   for polymerases during amplification reactions. It is therefore useful
CC   for monitoring nutrients or medicines and environmental factors such as
CC   radiation from televisions or computer screens. The method is applicable
CC   to all types of mutations (collectively or individually), is unaffected
CC   by repair mechanisms or in vivo selection mechanisms, and is an in vitro
CC   process so animal testing is not required. AA24098-224105 represent
CC   rat NF-1 DNA fragments used to illustrate the method of the invention.
XX
SQ   Sequence 240 BP; 66 A; 52 C; 54 G; 68 T; 0 other;

Query Match        8.7%;  Score 32.8;  DB 20;  Length 240;
Best Local Similarity  76.9%;  Pred. No. 1.7;
Matches  40;  Conservative   0;  Mismatches  12;  Indels   0;  Gaps   0;

Qy   155 AATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTT 206
         ||||| ||| ||  ||||||||||||||||||||||| | |||  ||| ||||
Db    53 AATCTGATCCTATATCTAAGAAAAAACCTCCGTGCCAAGTCGAAGTTGCTTTT 2

RESULT 14
AA224102/c
ID   AA224102 standard; DNA; 240 BP.
XX
AC   AA224102;
XX
DT   04-FEB-2000  (first entry)
XX
DE   Rat NF-1 DNA fragment 5.
XX
KW   NF-1; rat; primer; mutagenicity; mutation rate; polymerase error rate;
KW   radiation; ss.
XX
OS   Rattus sp.
XX
PN   DE19818422-A1.
XX
PD   28-OCT-1999.
XX
PF   24-APR-1998;  98DE-1018422.
XX
PR   24-APR-1998;  98DE-1018422.
XX
PA   (HARJ/) HARJES S.
XX
PI   Harjes S;
XX
DR   WPI; 1999-602407/52.
XX
PT   Method for determining absolute mutagenicity and error rate of
PT   polymerases in amplification reactions -
XX
PS   Examples; Fig 5; 18pp; German.
XX
CC   This invention describes a novel method for determining the absolute
CC   mutagenicity of a substance. The method is used to determine the
CC   mutation rate associated with particular substances and the error rates
CC   for polymerases during amplification reactions. It is therefore useful
CC   for monitoring nutrients or medicines and environmental factors such as
CC   radiation from televisions or computer screens. The method is applicable
CC   to all types of mutations (collectively or individually), is unaffected
CC   by repair mechanisms or in vivo selection mechanisms, and is an in vitro
CC   process so animal testing is not required. AA24098-224105 represent
CC   rat NF-1 DNA fragments used to illustrate the method of the invention.
XX
SQ   Sequence 240 BP; 66 A; 52 C; 54 G; 68 T; 0 other;

Query Match        8.7%;  Score 32.8;  DB 20;  Length 240;
Best Local Similarity  76.9%;  Pred. No. 1.7;
Matches  40;  Conservative   0;  Mismatches  12;  Indels   0;  Gaps   0;

Qy   155 AATCTAATGTTCTTTGGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTT 206
         ||||| ||| ||  ||||||||||||||||||||||| | |||  ||| ||||
Db    53 AATCTGATCCTATATCTAAGAAAAAACCTCCGTGCCAAGTCGAAGTTGCTTTT 2

RESULT 15
AA224103/c
ID   AA224103 standard; DNA; 240 BP.
XX
AC   AA224103;
XX
DT   04-FEB-2000  (first entry)
XX
DE   Rat NF-1 DNA fragment 6.
XX
KW   NF-1; rat; primer; mutagenicity; mutation rate; polymerase error rate;
KW   radiation; ss.
XX
OS   Rattus sp.
XX
PN   DE19818422-A1.
XX
PD   28-OCT-1999.
XX
PF   24-APR-1998;  98DE-1018422.
XX
PR   24-APR-1998;  98DE-1018422.
XX
PA   (HARJ/) HARJES S.
XX
PI   Harjes S;
XX
DR   WPI; 1999-602407/52.
XX
PT   Method for determining absolute mutagenicity and error rate of
PT   polymerases in amplification reactions -
XX
PS   Examples; Fig 5; 18pp; German.
XX
CC   This invention describes a novel method for determining the absolute
CC   mutagenicity of a substance. The method is used to determine the
CC   mutation rate associated with particular substances and the error rates
CC   for polymerases during amplification reactions. It is therefore useful
CC   for monitoring nutrients or medicines and environmental factors such as
CC   radiation from televisions or computer screens. The method is applicable
CC   to all types of mutations (collectively or individually), is unaffected
CC   by repair mechanisms or in vivo selection mechanisms, and is an in vitro
```

CC   process so animal testing is not required. AA224098-224105 represent
CC   rat NF-1 DNA fragments used to illustrate the method of the invention.
XX
SO   Sequence 240 BP; 66 A; 52 C; 54 G; 68 T; 0 other;

Query Match   8.7%;   Score 32.8; DB 20;   Length 240;
Best Local Similarity   76.9%;   Pred. No. 1.7;
Matches   40; Conservative   0; Mismatches   12; Indels   0; Gaps   0;

Qy   155 AATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTTACTTTGGTTTT 206
           ||||| ||| | || | |||||||||||| ||||||||| | | | ||| ||||
Db   53 AATCTGATGCTATTATCTAAGAAAAACCTCCGTGCCAAGTCGGAAGTTGCTTTT 2

Search completed: January 31, 2003, 01:17:24
Job time : 514 secs

us-09-803-719-222.rge

OM nucleic - nucleic search, using sw model

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: January 31, 2003, 01:15:03 ; Search time 2759 Seconds
                 (without alignments)
                 3987.264 Million cell updates/sec

Title: US-09-803-719-222
Perfect score: 378
Sequence: 1 tacggctgcttaagacgact..........taccattggagaaccatgcg 378
Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 1455140287B residues
Total number of hits satisfying chosen parameters: 4109280
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
    1: gb_ba:*
    2: gb_htg:*
    3: gb_in:*
    4: gb_om:*
    5: gb_ov:*
    6: gb_pat:*
    7: gb_ph:*
    8: gb_pl:*
    9: gb_pr:*
    10: gb_ro:*
    11: gb_sts:*
    12: gb_sy:*
    13: gb_un:*
    14: gb_vi:*
    15: em_ba:*
    16: em_fun:*
    17: em_hum:*
    18: em_in:*
    19: em_mu:*
    20: em_om:*
    21: em_or:*
    22: em_ov:*
    23: em_pat:*
    24: em_ph:*
    25: em_pl:*
    26: em_pro:*
    27: em_sts:*
    28: em_un:*
    29: em_vi:*
    30: em_htg_hum:*
    31: em_htg_inv:*
    32: em_htg_mus:*
    33: em_htg_other:*
    34: em_htg_pln:*
    35: em_htg_rod:*
    36: em_htg_mam:*
    37: em_htg_vrt:*
    38: em_sy:*
    39: em_htgo_hum:*
    40: em_htgo_mus:*
    41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	378	100.0	378	6	AX245292	AX245292 Sequence
2	299.6	79.3	327	6	AX245346	AX245346 Sequence
3	279.6	74.0	322	6	AX245365	AX245365 Sequence
4	275	72.8	318	6	AX245370	AX245370 Sequence
5	119	31.5	59231	9	AL135937	AL135937 Human DNA
6	72.8	17.9	162394	2	AC119775	AC119775 Rattus no
7	53.8	14.2	223604	10	AL713981	AL713981 Mouse DNA
8	41	10.8	176543	2	AL845312	AL845312 Danio rer
9	39.4	10.4	176030	2	AC129650	AC129650 Rattus no
10	39.2	10.4	91071	8	AC002339	AC002339 Arabidops
11	38.4	10.2	130206	2	AC104652	AC104652 Homo sapi
12	38.4	10.2	152822	9	AC108866	AC108866 Homo sapi
13	38.4	10.2	190912	3	AC080170	AC080170 Homo sapi
14	38.2	10.1	30960	3	AC024785	AC024785 Caenorhab
15	37.8	10.0	174196	2	AL732628	AL732628 Danio rer
16	37.8	10.0	177368	9	AC097813	AC097813 Rattus no
17	37.4	9.9	137193	2	AC110131	AC110131 Rattus no
18	37.4	9.9	160582	9	AC108063	AC108063 Homo sapi
19	37.4	9.9	192297	2	AC013528	AC013528 Homo sapi
20	37.4	9.9	259272	2	AC129172	AC129172 Rattus no
21	37.2	9.8	165118	2	AC096436	AC096436 Rattus no
22	37	9.8	1350	8	AC107535	AC107535 Rattus no
23	37	9.8	158968	10	U91924	U91924 Nicotiana t
24	36.4	9.6	1980	3	AC124196	AC124196 Mus muscu
25	36.2	9.6	50840	9	AL662785	AL662785 Mouse DNA
26	36.2	9.6	127508	9	U59830	U59830 Loligo opal
27	36.2	9.6	162921	9	AC093598	AC093598 Homo sapi
28	36.2	9.6	1733526	9	AP002964	AP002964 Homo sapi
29	36.2	9.6	165282	9	AP000822	AP000822 Human DNA
30	36	9.5	62254	2	AC100028	AC100028 Homo sapi
31	35.8	9.5	62582	2	AC103992	AC103992 Homo sapi
32	35.8	9.5	162605	2	AC068698	AC068698 Homo sapi
33	35.8	9.5	173760	2	AL354653	AL354653 Human DNA
34	35.8	9.5	188515	2	AC102531	AC102531 Mus muscu
35	35.8	9.5	210386	10	AL663065	AL663065 Mouse DNA
36	35.8	9.5	1110000	2	AC092450_2	Continuation (3 of
37	35.6	9.4	1117296	2	AC092865	AC092865 Homo sapi
38	35.6	9.4	155415	2	AC108886	AC108886 Rattus no
39	35.6	9.4	162173	8	AP002953	AP002953 Oryza sat
40	35.6	9.4	163317	8	AC095564	AC095564 Homo sapi
41	35.6	9.4	204632	2	AC097583	AC097583 Rattus no
42	35.6	9.4	163311	2	AB015609	AB015609 Nicotiana
43	35.4	9.4	1537	8	AF033100	AF033100 Nicotiana
44	35.4	9.4	2325	8	AB010426	AB010426 Phytoplas
45	35.4	9.4	7005	1		

ALIGNMENTS

RESULT 1
AX245292

LOCUS   AX245292   378 bp   DNA   linear   PAT 28-SEP-2001
DEFINITION Sequence 222 from Patent WO0166753.
ACCESSION   AX245292
VERSION   AX245292.1 GI:15859966
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 378)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
            Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,

```
TITLE       Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
             Stache-Crain,B.
JOURNAL     Human genes and gene expression products
             Patent: WO 0166753-A 222 13-SEP-2001;
             Chiron Corporation (US); Hyseq Inc. (US)
FEATURES             Location/Qualifiers
    source   1..378
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
BASE COUNT   100 a          76 c         101 g        101 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 2.3e-103;
Matches 378; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

                                       Score 378; DB 6; Length 378;

Qy      1 TACGGCTGCTTAAGACGACTTAAGGGGGAATGACGACCAGCGGCCTCTTAGAGGAACATATGG 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TACGGCTGCTTAAGACGACTTAAGGGGGAATGACGACCAGCGGCCTCTTAGAGGAACATATGG 60

Qy     61 AAAACACCCAAGCCGCGGAGAGTCTCTCACAAGCTTGAATGTGTGTTCTCGGAGCTGAAGGATGC 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAAACACCCAAGCCGCGGAGAGTCTCTCACAAGCTTGAATGTGTGTTCTCGGAGCTGAAGGATGC 120

Qy    121 ACGGTTGTTAAGCCCCTGTTCTTTCCGTTGTTTAATCTAATGTGTTCTTTTGAATAAAAAC 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACGGTTGTTAAGCCCCTGTTCTTTCCGTTGTTTAATCTAATGTGTTCTTTTGAATAAAAAC 180

Qy    181 CTCCCTGCCAAGTAGTACTTGGTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGACC 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTCCCTGCCAAGTAGTACTTGGTTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGACC 240

Qy    241 TTTGGCACACATTGAAGGGATGGGATGGAGATGCCAGCGCCGCTCATGTGTTGGCGCGAG 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTTGGCACACATTGAAGGGATGGGATGGAGATGCCAGCGCCGCTCATGTGTTGGCGCGAG 300

Qy    301 CTATGTTGGTAGCTATAGCAGAAAGTCTTCTTTGGCAAAAGATTCCTCCCGGGAAGGAAGGTA 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTATGTTGGTAGCTATAGCAGAAAGTCTTCTTTGGCAAAAGATTCCTCCCGGGAAGGAAGGTA 360

Qy    361 CCATTGGAGAACCATCCG 378
           ||||||||||||||||||
Db    361 CCATTGGAGAACCATGCG 378


RESULT 2
LOCUS       AX245346
DEFINITION  Sequence 276 from Patent WO0166753.
ACCESSION   AX245346
VERSION     AX245346.1  GI:15860020
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 327)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
            Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Kita,D.,
            Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
            Stache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0166753-A 276 13-SEP-2001;
            Chiron Corporation (US); Hyseq Inc. (US)
FEATURES             Location/Qualifiers
    source   1..327
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
BASE COUNT   92 a          64 c          81 g          90 t
ORIGIN

Query Match                                          79.3%; Score 299.6; DB 6; Length 327;
```

```
                                       Best Local Similarity 98.7%; Pred. No. 1.2e-79;
                                       Matches 302; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy     28 GAAGACGCAGCGCGGCGTCTCTTAGAGGAACATATGGAAAACACCCAAGCCGGAGAGTCTCTCACA 87
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GAATGACGCAGCAGCGTCTCGAGAGGAACATATGGAAAACACCCAAGCCGGAGAGTCTCTCACA 80

Qy     88 AGCTTGAAATGTGTGTTCTCTGGAGCTGAAGGATGCACGGTTGTTGTTAAGCCCCTGTTCTTTTCC 147
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 AGCTTGAAATGTGTGTTCTCTGGAGCTGAAGGATGCACGGTTGTTGTTAAGCCCCTGTTCTTTTCC 140

Qy    148 GTTGTTTAATCTAATGTGTTCTTTTGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA 207
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 GTTGTTTAATCTAATGTGTTCTTTTGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA 200

Qy    208 TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGGAT 267
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGGAT 260

Qy    268 GGAGATGCCAAATCCATGGAAATCAGGTGGGCGCGAGCTATGTTGGTAGCTATACCAGAAGTCT 327
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 GGAGATGCCAAATCCATGGAAATCAGGTGGGCCAGCGCTATGTTGGGTAGCTATAGCAGAAGTCT 320

Qy    328 TCTTGG 333
           ||||||
Db    321 TCTTGG 326


RESULT 3
LOCUS       AX245365
DEFINITION  Sequence 295 from Patent WO0166753.
ACCESSION   AX245365
VERSION     AX245365.1  GI:15860039
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 322)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
            Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Kennedy,G.C., Pot,D.,
            Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Dickson,M.,
            Stache-Crain,B.                      Kita,D., Garcia,V. and
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0166753-A 295 13-SEP-2001;
            Chiron Corporation (US); Hyseq Inc. (US)
FEATURES             Location/Qualifiers
    source   1..322
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
BASE COUNT   90 a          64 c          78 g          89 t         1 others
ORIGIN

Query Match                                          74.0%; Score 279.6; DB 6; Length 322;
Best Local Similarity 98.6%; Pred. No. 1.4e-73;
Matches 282; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy     28 GAATGACGCAGCGCGGCGTCTCTTAGAGGAACATATGGAAAACACCCAAGCCGGAGAGTCTCTCACA 87
           |||||||||| ||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db     36 GAATGACGCAGCAGCCTCGAGAGGAACATATGGAAAACACCCAAGCCGGAGAGTCTCTCACA 95

Qy     88 AGCTTGAATGTGTGTCTCGGAGCTGAAGGATGCACGGTTGTTGTTAAGCCCCTGTTCTTTTCC 147
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     96 AGCTTGAATGTGTGTTCTCGGAGCTGAAGGATGCACGGTTGTTGTTAAGCCCCTGTTCTTTTCC 155

Qy    148 GTTGTTTAATCTAATGTGTTCTTTTGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA 207
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    156 GTTGTTTAATCTAATGTGTTCTTTTGAATAAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA 215

Qy    208 TGCTCAACATGCTTTGACTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGGTGAT 267
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    216  TGCTCAACACATGCTTTGACTTGTTGAAAAGAGACCTTTGGCACACATTGAAGGATGGTGAT 275

RESULT 4
Qy    268  GGAGATGCCAATCCATGGAATCAGGTGCGCGACCTATGTTGGGTACC 313
           ||||||||||||||||||||||||||||||||| ||||||||||||||
Db    276  GGAGATGCCAATCCATGGAATCAGGTGGCACAGCTATGTGTGGTAGC 321

LOCUS       AX245370                 318 bp    DNA     linear   PAT 28-SEP-2001
DEFINITION  Sequence 300 from Patent WO0166753.
ACCESSION   AX245370
VERSION     AX245370.1  GI:15860044
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 318)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
            Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
            Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
            Stache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0166753-A 300 13-SEP-2001;
            Chiron Corporation (US); Hyseq Inc. (US)
FEATURES             Location/Qualifiers
     source          1..318
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       90 a      63 c      76 g      89 t
ORIGIN

Query Match             72.8%;  Score 275;  DB 6;  Length 318;
Best Local Similarity   98.2%;  Pred. No. 3.4e-72;
Matches  278;  Conservative  0;  Mismatches  5;  Indels  0;  Gaps  0;

Qy    28   GAATGACGCAGCCGGCCTCTTAGAGAGAACATATGAAAACACCCAAGCCGGAGTCTCTCACA 87
           ||||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||
Db    36   GAATGACGCAGCCAGCTCTCGAGAGGAACATAAGGAAAAACACCCAAGCCGAGTCTCTCACA 95

Qy    88   AGCTTGAATGTGTGTCTTCTCGGAGCTGAAGGATGCACGGTTGTGTTAAGCCCCTGTTCTTCTTTTCC 147
           |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    96   AGCTTGAATGTGTGTCTTCTCGGAGCTGAAGGATGCACGGTTGTGTTAAGCCCCTGTTCTTCTTTTCC 155

Qy    148  GTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCCAAGTAGTACTTGGTTTTA 207
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    156  GTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCGCCCAAGTAGTACTTGGTTTTA 215

Qy    208  TGCTCAACACATGCTTTGACTTGTTGAAAAGAGACCTTTGGCACACATTGAAGGGATGGTGAT 267
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    216  TGCTCAACACATGCTTTGACTTGTTGAAAAGAGACCTTTGGCACACATTGAAGGATGGTGAT 275

Qy    268  GGAGATGCCAATCCATGGAATCAGGTGCGCGACCTATGTTGGT 310
           |||||||||||||||||||||||||||||||| |||||||||||
Db    276  GGAGATGCCAATCCATGGAATCAGGTGGCACAGCTATGTGTGGT 318

RESULT 5
LOCUS       AL135937               59231 bp    DNA     linear   PRI 15-MAR-2001
DEFINITION  Human DNA sequence from clone RP1-278022 on chromosome 20. Contains
            STSs and GSSs, complete sequence.
ACCESSION   AL135937
VERSION     AL135937.22  GI:7248202
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 59231)
  AUTHORS   Phillimore,B.
```

REFERENCE section (right column):
```
            Direct Submission
  TITLE     Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
  JOURNAL   CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 15, 2000 this sequence version replaced gi:7242336.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            IMPORTANT: This sequence is not the entire insert of clone
            RP1-278022 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP5-1099D15 is at 59132 in this sequence.
            The true left end of clone RP4-697P8 is at 100 in this sequence.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP1-278022 is from
            the library RPCI-1 constructed by the group of Pieter de Jong. For
            further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
FEATURES             Location/Qualifiers
     source          1..59231
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP1-278022"
                     /clone_lib="RPCI-1"
     repeat_region   58..213
                     /note="MER5A repeat: matches 9..188 of consensus"
     repeat_region   233..409
                     /note="MER3 repeat: matches 6..191 of consensus"
     repeat_region   572..795
                     /note="MIR repeat: matches 3..241 of consensus"
     repeat_region   2927..2992
                     /note="Alu repeat: matches 242..307 of consensus"
     repeat_region   3019..3635
                     /note="L1MC4 repeat: matches 7117..7739 of consensus"
     repeat_region   3880..4411
                     /note="L1ME2 repeat: matches 5541..6154 of consensus"
     repeat_region   4412..4600
                     /note="L1M4 repeat: matches 4015..4580 of consensus"
     misc_feature    complement(5199..5709)
                     /note="MER58A repeat: matches 32..224 of consensus"
     repeat_region   7115..7585
                     /note="match: GSS: Em:AQ613263"
     repeat_region   7631..7695
                     /note="Charlie4a repeat: matches 25..507 of consensus"
     repeat_region   8546..9223
                     /note="MIR repeat: matches 13..81 of consensus"
     repeat_region   9986..10209
                     /note="L1MC4 repeat: matches 1..225 of consensus"
     repeat_region   10225..10250
                     /note="AluSq repeat: matches 1..225 of consensus"
     misc_feature    complement(10684..11081)
                     /note="13 copies 2 mer aa 96% conserved"
     repeat_region   11093..11168
                     /note="match: GSS: Em:AQ662878"
                     /note="38 copies 2 mer aa 72% conserved"
```

```
repeat_region   11575..12292
                /note="AluSg repeat: matches 5545..6278 of consensus"
misc_feature    12860..13177
                /note="match: GSS: Em:AQ035618"
repeat_region   13327..13423
                /note="L2 repeat: matches 2657..2750 of consensus"
repeat_region   13719..13813
                /note="Aluo repeat: matches 1..293 of consensus"
repeat_region   13814..14087
                /note="L2 repeat: matches 2195..2293 of consensus"
repeat_region   14088..14219
                /note="L1MA5 repeat: matches 6025..6298 of consensus"
repeat_region   14965..15067
                /note="L2 repeat: matches 2078..2195 of consensus"
misc_feature    15849..16248
                /note="match: GSS: Em:AQ754141"
misc_feature    complement(15295..15803)
                /note="MIR repeat: matches 169..262 of consensus"
repeat_region   15961..16498
                /note="match: GSS: Em:AQ694331"
repeat_region   16499..16808
                /note="L1ME3A repeat: matches 5427..5964 of consensus"
repeat_region   16809..17004
                /note="AluSx repeat: matches 1..310 of consensus"
repeat_region   17356..17544
                /note="L1ME3A repeat: matches 5964..6159 of consensus"
repeat_region   18267..18357
                /note="MIR repeat: matches 26..230 of consensus"
repeat_region   18358..18690
                /note="L1Md repeat: matches 50..139 of consensus"
repeat_region   18690..18717
                /note="MIR repeat: matches 17..410 of consensus"
repeat_region   18691..18717
                /note="MER58B repeat: matches 2..341 of consensus"
repeat_region   18717..18885
                /note="MIR repeat: matches 24..50 of consensus"
repeat_region   18885..18952
                /note="AluSg repeat: matches 1..288 of consensus"
misc_feature    complement(19364..19600)
                /note="34 copies 2 mer ga 79% conserved"
repeat_region   22132..22242
                /note="LTR16C repeat: matches 1..387 of consensus"
repeat_region   22243..22551
                /note="LTR33 repeat: matches 410..521 of consensus"
repeat_region   22552..22951
                /note="AluSx repeat: matches 1..310 of consensus"
repeat_region   23356..23643
                /note="LTR33 repeat: matches 17..410 of consensus"
repeat_region   24124..24510
                /note="AluSg repeat: matches 1..288 of consensus"
repeat_region   24747..24810
                /note="LTR16C repeat: matches 1..387 of consensus"
repeat_region   24967..25054
                /note="32 copies 2 mer ta 68% conserved"
repeat_region   25134..25443
                /note="L1M1 repeat: matches 5678..5763 of consensus"
repeat_region   25767..26069
                /note="AluSg repeat: matches 1..313 of consensus"
repeat_region   26295..26579
                /note="AluSg repeat: matches 1..304 of consensus"
repeat_region   27383..27481
                /note="L1MC/D repeat: matches 5290..5579 of consensus"
repeat_region   28338..28469
                /note="MIR repeat: matches 20..142 of consensus"
repeat_region   28472..28583
                /note="L1MC4 repeat: matches 7841..7973 of consensus"
repeat_region   28584..28970
                /note="56 copies 2 mer ta 73% conserved"
repeat_region   30290..30876
                /note="L1MC4 repeat: matches 7430..7849 of consensus"
repeat_region   30877..31223
                /note="L2 repeat: matches 2124..2658 of consensus"
repeat_region   31224..31421
                /note="MER1B repeat: matches 1..337 of consensus"
repeat_region   31422..31739
                /note="L2 repeat: matches 1945..2124 of consensus"
repeat_region   31750..32046
                /note="AluY repeat: matches 1..309 of consensus"

repeat_region   32047..32188
                /note="AluSg repeat: matches 3..298 of consensus"
repeat_region   32189..32474
                /note="L2 repeat: matches 1809..1937 of consensus"
repeat_region   32475..32713
                /note="Aluo repeat: matches 1..293 of consensus"
repeat_region   34869..34938
                /note="L2 repeat: matches 1420..1809 of consensus"
repeat_region   35160..35351
                /note="MIR repeat: matches 191..261 of consensus"
repeat_region   36121..36273
                /note="MER5A repeat: matches 7..189 of consensus"
repeat_region   37904..37941
                /note="MIR repeat: matches 66..225 of consensus"
repeat_region   38788..38840
                /note="19 copies 2 mer aa 81% conserved"
repeat_region   38841..39254
                /note="MSTB repeat: matches 1..426 of consensus"
repeat_region   39255..39304
                /note="MIR repeat: matches 84..150 of consensus"
repeat_region   39310..39339
                /note="L1Md repeat: matches 251..280 of consensus"
repeat_region   39314..39717
                /note="L1M3e repeat: matches -879..779 of consensus"
repeat_region   39724..39990
                /note="L1MA9 repeat: matches 3735..4011 of consensus"
repeat_region   39992..40126
                /note="Aluo repeat: matches 2..134 of consensus"
repeat_region   40153..40379
                /note="L1MA9 repeat: matches 3999..4226 of consensus"
repeat_region   40380..40675
                /note="Aluo repeat: matches 1..306 of consensus"
repeat_region   40676..42726
                /note="L1MA9 repeat: matches 4226..6308 of consensus"
repeat_region   43333..43524
                /note="L1MA9 repeat: matches 165..386 of consensus"
misc_feature    complement(43484..44004)
                /note="LTR16C repeat: matches 165..386 of consensus"
repeat_region   44765..44818
                /note="match: GSS: Em:AQ800109"
misc_feature    complement(45503..46337)
                /note="MIR repeat: matches 200..250 of consensus"
```

RESULT 6
AC119775/c

```
LOCUS       AC119775      162394 bp    DNA     linear    HTG 18-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-464J18, *** SEQUENCING IN PROGRESS
            ***, 61 unordered pieces.
ACCESSION   AC119775
VERSION     AC119775.4  GI:21747184
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus
ORGANISM    Rattus norvegicus
            Norway rat.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 162394)
AUTHORS     Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
```

```
Query Match      31.5%;  Score 119; DB 9; Length 59231;
Best Local Similarity   100.0%;
Matches 119;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    148 GTTGTTTAATCTAAATGTTCTTTTGGAATAAAAAACCTCCCTGCCCAAGTAGTACTTGGTTTTA 207
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18738 GTTGTTTAATCTAAATGTTCTTTTGGAATAAAAAACCTCCCTGCCCAAGTAGTACTTGGTTTTA 18797

Qy    208 TGCTCAACATGCTTTGACTGTTGAAAAAGAGAACCTTTGGCACACATTGAAGGGGATGGGTGA 266
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18798 TGCTCAACATGCTTTGACTGTTGAAAAAGAGAACCTTTGGCACACATTGAAGGGGATGGGTGA 18856
```

Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J, Draper,H. Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A. Hogues,M. Holloway,C., Hollins,B.,
Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B. Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E. Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newtson,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,H., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 162394)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (02-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 162394)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Jul 14, 2002 this sequence version replaced gi:20467838.
COMMENT   Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu

            Project Information
              Center project name: GVQM
              Center clone name: CH230-464J18

            Summary Statistics
              Sequencing vector: plasmid;
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 104933 bases at least Q40
              Consensus quality: 112584 bases at least Q30
              Consensus quality: 117408 bases at least Q20

            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 61 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *   1   1004: contig of 1004 bp in length
            *   1005   1104: gap of unknown length
            *   1105   2121: contig of 1017 bp in length
            *   2122   2221: gap of unknown length
            *   2222   3241: contig of 1020 bp in length
            *   3242   3341: gap of unknown length
            *   3342   4689: contig of 1348 bp in length
            *   4690   4789: gap of unknown length
            *   4790   6424: contig of 1635 bp in length
            *   6425   6524: gap of unknown length
            *   6525   8023: contig of 1499 bp in length
            *   8024   8123: gap of unknown length
            *   8124   9618: contig of 1495 bp in length
            *   9619   9718: gap of unknown length
            *   9719   10759: contig of 1041 bp in length
            *   10760   10859: gap of unknown length
            *   10860   12638: contig of 1779 bp in length
            *   12639   12738: gap of unknown length
            *   12739   14083: contig of 1345 bp in length
            *   14084   14183: gap of unknown length
            *   14184   15891: contig of 1708 bp in length
            *   15892   15991: gap of unknown length
            *   15992   17623: contig of 1632 bp in length
            *   17624   17723: gap of unknown length
            *   17724   18962: contig of 1239 bp in length
            *   18963   19062: gap of unknown length
            *   19063   20849: contig of 1787 bp in length
            *   20850   20949: gap of unknown length
            *   20950   22169: contig of 1220 bp in length
            *   22170   22269: gap of unknown length
            *   22270   23483: contig of 1214 bp in length
            *   23484   23583: gap of unknown length
            *   23584   24875: contig of 1292 bp in length
            *   24876   24975: gap of unknown length
            *   24976   26670: contig of 1695 bp in length
            *   26671   26770: gap of unknown length
            *   26771   27914: contig of 1144 bp in length
            *   27915   28014: gap of unknown length
            *   28015   29124: contig of 1110 bp in length
            *   29125   29224: gap of unknown length
            *   29225   31386: contig of 2162 bp in length
            *   31387   31486: gap of unknown length
            *   31487   32500: contig of 1014 bp in length
            *   32501   32600: gap of unknown length
            *   32601   34345: contig of 1745 bp in length
            *   34346   34445: gap of unknown length
            *   34446   36597: contig of 2152 bp in length
            *   36598   36697: gap of unknown length
            *   36698   38768: contig of 2071 bp in length
            *   38769   38868: gap of unknown length
            *   38869   40546: contig of 1678 bp in length
            *   40547   40646: gap of unknown length
            *   40647   41770: contig of 1124 bp in length
            *   41771   41870: gap of unknown length
            *   41871   43735: contig of 1865 bp in length
            *   43736   43835: gap of unknown length
            *   43836   45206: contig of 1371 bp in length
            *   45207   45306: gap of unknown length
            *   45307   47814: contig of 2508 bp in length
            *   47815   47914: gap of unknown length
            *   47915   49406: contig of 1492 bp in length
            *   49407   49506: gap of unknown length
            *   49507   51910: contig of 2404 bp in length
            *   51911   52010: gap of unknown length
            *   52011   54105: contig of 2095 bp in length
            *   54106   54205: gap of unknown length
            *   54206   56954: contig of 2749 bp in length
            *   56955   57054: gap of unknown length

RESULT 7

```
 *  57055 59393: contig of 2339 bp in length
 *  59394 59493: gap of unknown length
 *  59494 61754: contig of 2261 bp in length
 *  61754 61854: gap of unknown length
 *  61755 64055: contig of 2201 bp in length
 *  61855 64055: gap of unknown length
 *  64056 64155: gap of unknown length
 *  64056 67129: contig of 2974 bp in length
 *  64156 67129: gap of unknown length
 *  67130 67229: gap of unknown length
 *  67230 69622: contig of 2393 bp in length
 *  67230 69622: gap of unknown length
 *  69623 69722: gap of unknown length
 *  69723 72435: contig of 2713 bp in length
 *  69723 72435: gap of unknown length
 *  72436 72535: gap of unknown length
 *  72436 75742: contig of 3207 bp in length
 *  72536 75742: gap of unknown length
 *  75743 75842: gap of unknown length
 *  75743 77891: contig of 2049 bp in length
 *  75843 77891: gap of unknown length
 *  77892 77991: gap of unknown length
 *  77892 80486: contig of 2495 bp in length
 *  77992 80486: gap of unknown length
 *  80487 80586: gap of unknown length
 *  80487 83986: contig of 3400 bp in length
 *  80587 83986: gap of unknown length
 *  83987 84086: gap of unknown length
 *  83987 86359: contig of 2273 bp in length
 *  84087 86359: gap of unknown length
 *  86360 86459: gap of unknown length
 *  86360 89249: contig of 2790 bp in length
 *  86460 89249: gap of unknown length
 *  89250 89349: gap of unknown length
 *  89250 91349: contig of 2000 bp in length
 *  89350 91349: gap of unknown length
 *  91350 91449: gap of unknown length
 *  91350 94806: contig of 3457 bp in length
 *  91450 94806: gap of unknown length
 *  94807 94906: gap of unknown length
 *  94807 98893: contig of 4088 bp in length
 *  94907 98893: gap of unknown length
 *  98894 98993: gap of unknown length
 *  98994 105047: contig of 5953 bp in length
 *  99095 105047: gap of unknown length
 * 105048 105147: gap of unknown length
 * 105048 108667: contig of 3521 bp in length
 * 105148 108667: gap of unknown length
 * 108668 108767: gap of unknown length
 * 108669 111828: contig of 3061 bp in length
 * 108769 111828: gap of unknown length
 * 111829 111928: gap of unknown length
 * 111830 116396: contig of 4468 bp in length
 * 111929 116396: gap of unknown length
 * 116397 116496: gap of unknown length
 * 116398 121200: contig of 4790 bp in length
 * 116497 121200: gap of unknown length
 * 121201 121300: gap of unknown length
 * 121202 116804: contig of 5604 bp in length
 * 122102 122201: gap of unknown length
 * 122202 126351: contig of 4151 bp in length
 * 122202 126352: gap of unknown length
```

```
Query Match    17.9%;     Score 67.8;  DB 2;  Length 162394;
Best Local Similarity  73.1%;  Pred. No. 1.7e-09;
Matches   87;  Conservative  0;  Mismatches 32;  Indels  0;  Gaps  0;

Qy   148 GTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA 207
          || ||| |||||| |||| ||||| ||||| ||||||||||||||||||||| | |||
Db 79083 GTGGTTTATTCTAAAGTTCCCTGAAATTAAAACCTCCCTGCAACTCTAAGACTTGGATTA 79024

Qy   208 TGCTCAACATGCTTTGACTGTTGAAAAAGAGACCATTGAAGGGATGGTGA 266
          |||||||||| |||||||||||| | ||||||||||||||||| |||
Db 79023 ATCTCAACATGCTTTGACTGTTAAAAAGAGGCCCTTTGATCACACTAGAGGGACACTAA 78965
```

LOCUS　　　　AL713981

DEFINITION　Mouse DNA sequence from clone RP23-173012 on chromosome 2, complete sequence.　　　223604 bp　DNA　linear　ROD 28-JUN-2002

ACCESSION　AL713981

VERSION　　AL713981.16　GI:21627973

KEYWORDS　HTG.

SOURCE　　house mouse.

ORGANISM　Mus musculus
　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE　1　(bases 1 to 223604)

AUTHORS　Dunn,M.

TITLE　Direct Submission

JOURNAL　Submitted (27-JUN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT　On Jun 28, 2002 this sequence version replaced gi:21531387.

Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source sequence databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-173012 is
from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

FEATURES　　　Location/Qualifiers
　source　　　1..223604
　　　/organism="Mus musculus"
　　　/db_xref="taxon:10090"
　　　/chromosome="2"
　　　/clone="RP23-173012"
　　　/clone_lib="RPCI-23"

BASE COUNT　62140 a　46575 c　46966 g　67923 t

ORIGIN

RESULT 8

```
Qy   148 GTTGTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTA 207
          || ||| |||||| |||| ||||| ||||| |||||||||||||||||||| | |||
Db 178805 GTGGTTTATTCTAAAGTTCCCTGAAATTAAAACCTCCCTGCAACTCTAAGACTTGGACTA 178864

Qy   208 TGCTCAACATGCTTTGACTGTTGAAAAAGAGACCATTGAAGGGATGGTGA 266
          |||||||||| |||||||||||| ||||| ||||||||||| ||||
Db 178865 ATCTCAGCATGTTCTGACTGTTG--AACAGACCTTTGACACACAGTACAGGGGATAGTAA 178921
```

```
Query Match    14.2%;     Score 53.8;  DB 10;  Length 223604;
Best Local Similarity  71.4%;  Pred. No. 2.9e-05;
Matches   85;  Conservative  0;  Mismatches 32;  Indels  2;  Gaps  1;
```

LOCUS　　　　AL845312/c

DEFINITION　Danio rerio clone DKEY-287F10, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.　　　176543 bp　DNA　linear　HTG 09-AUG-2002

ACCESSION　AL845312

VERSION　　AL845312.1　GI:22205007

KEYWORDS　HTG; HTGS_PHASE1.

SOURCE　　zebrafish.

ORGANISM　Danio rerio
　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

REFERENCE　1　(bases 1 to 176543)

AUTHORS　Burton,J.

TITLE　Direct Submission

JOURNAL　Submitted (04-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfacc@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT
　-----------　Genome Center
　Center: Wellcome Trust Sanger Institute
　Center code: SC
　Web site: http://www.sanger.ac.uk

```
Contact: zface@sanger.ac.uk
--- Project Information
Center project name: zK287F10
--- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator
Consensus quality: 175506 bases at least Q40
Consensus quality: 175799 bases at least Q30
Consensus quality: 176033 bases at least Q20
Consensus quality: sum-of-contigs
Insert size: 176343; sum-of-contigs
Insert size: 177563; 4.7% error; agarose-fp
Quality coverage: 10.13x in Q20 bases; agarose-fp
coverage: 10.57x in Q20 bases; agarose-fp
.
.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
.
*   1   70191: contig of 70191 bp in length
*   70192 70291: gap of          100 bp
*   70292 166519: contig of 96228 bp in length
*   166520 166619: gap of        100 bp
*   166620 175543: contig of 9924 bp in length.

FEATURES             Location/Qualifiers
     source          1..175543
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="DKEY-287F10"
                     /clone_lib="DanioKey"
     misc_feature    1..70191
                     /note="assembly_fragment:03314
                     fragment_chain:1"
     misc_feature    70292..166519
                     /note="assembly_fragment:02085
                     fragment_chain:1"
     misc_feature    166620..175543
                     /note="assembly_fragment:01906.0"
BASE COUNT    57205 a  31955 c  31568 g  55615 t   200 others
ORIGIN
```

Query Match   10.8%;   Score 41; DB 2;   Length 176543;
Best Local Similarity   56.2%;   Pred. No. 0.21;
Matches   77; Conservative   0; Mismatches 60; Indels   0; Gaps   0;

```
Qy   200  TGGTTTTATGCTCAACATCGTTTGACTGTTGAAAAGAGACCTTGGCACACAATTGAAGG 259
          |||  || ||||||| ||||||||||||||||| |||||||||| ||||| |||||||||
Db  20135  TGGACGTGTGTTCAACATGCAATAAAATGTGGATAAGACCAATGAAGCACTTTTGAAGGC 20076

Qy   260  ATGGTGATGGAGATCCCAATCCATGGAATCAGGTGGCGCAGCTATGTTGGTAGCTATAGC 319
          ||| ||||||||||||| ||||||||| |||| ||  ||||||| ||||||| || ||||
Db  20075  ATGTTTCAGTATAAAACAATTAGTCGCATCACAAGGCTCTCCAAAGTTTCATGCAATTTC 20016

Qy   320  AGAAGTCTGTCTTGGCAA 336
          ||  | |||  |||||||
Db  20015  GGGTGTTTCATTTTTAA 19999
```

```
RESULT 9
ACI29650
LOCUS          ACI29650              176030 bp    DNA     linear    HTG 31-JUL-2002
DEFINITION     Rattus norvegicus clone CH230-9H1, *** SEQUENCING IN PROGRESS ***,
               67 unordered pieces.
ACCESSION      ACI29650
VERSION        ACI29650.1  GI:22024419
KEYWORDS       HTG; HTGS_PHASE1.
SOURCE         Norway rat.
  ORGANISM     Rattus norvegicus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
```

Rattus.
REFERENCE   1 (bases 1 to 176030)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Deerinck,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.H., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,J., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,R., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,D., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wieczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 176030)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Direct Submission
            Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA

```
COMMENT         --------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
                --------------- Project Information
                Center project name: GDLF
                Center clone name: CH230-9H1
                --------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 115101 bases at least Q40
                Consensus quality: 115101 bases at least Q30
                Consensus quality: 120951 bases at least Q20
                Consensus quality: 125022 bases at least Q20
                ---------------
*  NOTE: Estimated insert size may differ from sequence length
*  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank.draft.data.html).
*  NOTE: This is a 'working draft' sequence. It currently
*  consists of 67 contigs. The true order of the pieces
*  is not known and their order in this sequence record is
*  arbitrary. Gaps between the contigs are represented as
*  runs of N, but the exact sizes of the gaps are unknown.
```

```
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*        1  1355: contig of 1355 bp in length        *    60631  62480: contig of 1850 bp in length
*     1356  1455: gap of unknown length              *    62481  62580: gap of unknown length
*     1456  2495: contig of 1040 bp in length        *    62581  64336: contig of 1756 bp in length
*     2496  2595: gap of unknown length              *    64337  64436: gap of unknown length
*     2596  3722: contig of 1127 bp in length        *    64437  66314: contig of 1878 bp in length
*     3723  3822: gap of unknown length              *    66315  66414: gap of unknown length
*     3823  5539: contig of 1717 bp in length        *    66415  68864: contig of 2450 bp in length
*     5540  5639: gap of unknown length              *    68865  68964: gap of unknown length
*     5640  6946: contig of 1307 bp in length        *    68965  72244: contig of 3280 bp in length
*     6947  7046: gap of unknown length              *    72245  72344: gap of unknown length
*     7047  8274: contig of 1228 bp in length        *    72345  76090: contig of 3746 bp in length
*     8275  8374: gap of unknown length              *    76091  76190: gap of unknown length
*     8375  9445: contig of 1071 bp in length        *    76191  79415: contig of 3225 bp in length
*     9446  9545: gap of unknown length              *    79416  79515: gap of unknown length
*     9546 10624: contig of 1079 bp in length        *    79516  81318: contig of 1803 bp in length
*    10625 10724: gap of unknown length              *    81319  81418: gap of unknown length
*    10725 12447: contig of 1723 bp in length        *    81419  83055: contig of 1637 bp in length
*    12448 12547: gap of unknown length              *    83056  83155: gap of unknown length
*    12548 13952: contig of 1405 bp in length        *    83156  84871: contig of 1716 bp in length
*    13953 14052: gap of unknown length              *    84872  84971: gap of unknown length
*    14053 15149: contig of 1097 bp in length        *    84972  86611: contig of 1640 bp in length
*    15150 15249: gap of unknown length              *    86612  86711: gap of unknown length
*    15250 16707: contig of 1458 bp in length        *    86712  88684: contig of 1973 bp in length
*    16708 16807: gap of unknown length              *    88685  88784: gap of unknown length
*    16808 18601: contig of 1794 bp in length        *    88785  91052: contig of 2268 bp in length
*    18602 18701: gap of unknown length              *    91053  91152: gap of unknown length
*    18702 19807: contig of 1106 bp in length        *    91153  93547: contig of 2395 bp in length
*    19808 19907: gap of unknown length              *    93548  93647: gap of unknown length
*    19908 21089: contig of 1182 bp in length        *    93648  96794: contig of 3147 bp in length
*    21090 21189: gap of unknown length              *    96795  96894: gap of unknown length
*    21190 22504: contig of 1315 bp in length        *    96895 100411: contig of 3517 bp in length
*    22505 22604: gap of unknown length              *   100412 100511: gap of unknown length
*    22605 23644: contig of 1040 bp in length        *   100512 103952: contig of 3441 bp in length
*    23645 23744: gap of unknown length              *   103953 104052: gap of unknown length
*    23745 25245: contig of 1501 bp in length        *   104053 107723: contig of 3671 bp in length
*    25246 25345: gap of unknown length              *   107724 107823: gap of unknown length
*    25346 26792: contig of 1447 bp in length        *   107824 110233: contig of 2410 bp in length
*    26793 26892: gap of unknown length              *   110234 110333: gap of unknown length
*    26893 28163: contig of 1271 bp in length        *   110334 113225: contig of 2892 bp in length
*    28164 28263: gap of unknown length              *   113226 113325: gap of unknown length
*    28264 29783: contig of 1520 bp in length        *   113326 117459: contig of 4134 bp in length
*    29784 29883: gap of unknown length              *   117460 117559: gap of unknown length
*    29884 31670: contig of 1787 bp in length        *   117560 120591: contig of 3032 bp in length
*    31671 31770: gap of unknown length              *   120592 120691: gap of unknown length
*    31771 33311: contig of 1541 bp in length
*    33312 33411: gap of unknown length
*    33412 35781: contig of 2370 bp in length
*    35782 35881: gap of unknown length
*    35882 37513: contig of 1632 bp in length
*    37514 37613: gap of unknown length
*    37614 39621: contig of 2008 bp in length
*    39622 39721: gap of unknown length
*    39722 42050: contig of 2329 bp in length
*    42051 42150: gap of unknown length
*    42151 43951: contig of 1801 bp in length
*    43952 44051: gap of unknown length
*    44052 47095: contig of 3044 bp in length
*    47096 47195: gap of unknown length
*    47196 50224: contig of 3029 bp in length
*    50225 50324: gap of unknown length
*    50325 52186: contig of 1862 bp in length
*    52187 52286: gap of unknown length
*    52287 53628: contig of 1342 bp in length
*    53629 53728: gap of unknown length
*    53729 56328: contig of 2600 bp in length
*    56329 56428: gap of unknown length
*    56429 58107: contig of 1679 bp in length
*    58108 58207: gap of unknown length
*    58208 60530: contig of 2323 bp in length
*    60531 60630: gap of unknown length
```

```
RESULT 10
AC002339
LOCUS       AC002339               91071 bp    DNA     linear   PLN 11-MAR-2002
DEFINITION  Arabidopsis thaliana chromosome 2 clone T11A7 map CIC11C08,
            complete sequence.
ACCESSION   AC002339
VERSION     AC002339.3  GI:20196889
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Query Match             10.4%;  Score 39.4;  DB 2;  Length 176030;
Best Local Similarity   51.4%;
Matches   91;  Conservative   0;  Mismatches   86;  Indels   0;  Gaps   0;

Qy    127 GTTAAGCCCCTGTTCTTTTCCGTTGTTGTTTAATTCAAATGTTGTTCTTTTGGAATAAAAACCTCCT 186
          || |||  ||| ||| | || ||| || | ||  | ||||| ||||| |||| | |||  ||| ||
Db 114244 GATCTGTCCCATGCGTCTGATCAGTGGGTTGTTCCAGACTTCTCTGTTATGGACTTGCTCTCT 114303

Qy    187 GCCAAGTAGTACTTGGTTTTATGCTCAACATCTCGTTTGACTGTTTGAAAAGAGACCTTTGGC 246
          ||  || || || ||| ||| | ||| || || | || ||| | |||| |||| || || |||
Db 114304 TTATACTAATACGCCATGTTCCCCCACCAACCTCTCCTTTAATGAAAAGAAAGGGTGTTGCC 114363

Qy    247 ACACATTGAAAGGAATGGTGATGGAGAATGCCAATCCATGGAATCAGGTGCGGCGCAGCCTA 303
          || | | |  || ||   |  |   |||  || | ||| ||| | ||||  | || | || |
Db 114364 GCAGGTTGTTACGGACTCCATTTTGTTCCTATTGCAAGGTAGCCGATGCCAGAGGTA 114420
```

```
       Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
       Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 91071)
  AUTHORS   Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C.,
            Spriggs,T.A., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
            Somerville,C.R. and Venter,J.C.
  TITLE     Unpublished
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 91071)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 91071)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Apr 18, 2002 this sequence version replaced gi:6598350.
FEATURES             Location/Qualifiers
     source          1..91071
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /map="CIC11C08"
                     /clone="T11A7"
     misc_feature    1..13047
                     /note"overlap with BAC clone T6D20
                     (U90439:82778..95824)."
     gene            184..2478
                     /gene="At2g41890"
     mRNA            <184..>2478
                     /gene="At2g41890"
     CDS             184..2478
                     /gene="At2g41890"
                     /codon_start=1
                     /product="putative receptor-like protein kinase"
                     /protein_id="AAM14823.1"
                     /db_xref="GI:20196890"
                     /translation="MKMLRALLCLSLVFFLAFQIVVSEIQLGSKLVVGENTLWVSNN
                     GDFALGFFNPPGLLNRFSIGTWFNSNSIPIPDQRKVVWVAVAGAGVVVSDNSSYFELTRNG
                     ELVLFDSLLGVPVWNSKTNRFSVSSALLRDDGNILLLKDREIIVWQSFGTPTDTLLPN
                     QKFPAREMLRAASENSRSSYSLHLEDSGRLERLWRESNITEWSSGNEVVKKKRKKKKNI
                     GAVLTSEGALFLEDQDLMRPVWSVEGEDHNDTVKRFRLRLDRDGNLRMYSWNEDSRIW
                     KPWQAVENGCRVFAICSGSQVCEFNSSGYTECNCPENAFVAVSSVDPKCLVPYQKPGCKS
                     GFNMVKFKNLELYGIYPANDSVISQISSQRCKKLCLENSACPAVTYTNDGEPQCRMKL
                     TRYISGYSDPSLSSISYVKTCLDPIAVDPNNVSKESPVTVIKSHSICIPCLVGATSTT
                     LVLFLGFQLGTVVYIYRRKKLKLAKKAREFSKATNPRKGVMIFSWKTKAMTDNEDNNI
                     GPQIEKGVMPENELVAVKEVATLTEERKFRSSASKIGTMHIKNLANLEGYCCELGRR
                     FLVYEYAKNGSILDHIVDPLRSKLKTWRIRPDTCLSVAKALCYLHMECREFVSHGNLN
                     CGNILLGEDEAKLTEYGFGLCAADKGVEEILERVLRISFWCVQTDERLRPSMGEVVKVLEGTLSVDPP
                     GGRKETVVDKGLEGCFDVEEILERVLRISFWCVQTDERLRPSMGEVVKVLEGTLSVDPP
                     PPPACARSPPTNSSESSOSIYEP"
     repeat_region   810..833
                     /rpt_family="(GAA)n"
     gene            complement(3329..5857)
                     /gene="At2g41880"
                     /note="T11A7.23; supported by cDNA: gi_7861794_gb_AF204675
                     1_AF204675"
     mRNA            complement(join(<3329..3541,3637..3727,3807..3852,
                     3927..4062,4153..4274,4362..4724,4801..4930,5029..5105,
                     5183..5296,5446..5857))
                     /gene="At2g41880"
     CDS             complement(join(3460..3541,3637..3727,3807..3852,
                     3927..4062,4153..4274,4362..4724,4801..4930,5029..5105,
                     5183..5296,5446..5448))
                     /gene="At2g41880"
                     /codon_start=1
                     /product="putative guanylate kinase"
                     /protein_id="AAM14825.1"
                     /db_xref="GI:20196895"
                     /translation="MGEAPAVLVDHPENGHSNGVCVKSEPENTEITVDVGDRIFLIGG
```

```
                     /gene="At2g41870"
                     /note="T11A7.3; supported by cDNA: gi_11908071_gb_AF326883
                     1_AF326883"
THERNRNFSIGVQIYDKISNNWRFSPIVLGTGKPKSPGKGYSAFVLEQGRIIVKKGSPRND
SIWFLEVDSPYVREQKKLRKEVVAWSKGYRGNAEKPIVISGPSGVGKGTLISMLMKE
FPSMFGFSVSVHTTRSPRSMENDGVHHFADKKVMEKRIKDGKFLEFASVHGNLYGTSI
ESVEAVTDSGRCILDIDVQGARSVRASSLDAIFIFYCPPSNKELEDRLRARGTETEE
OIQKRLRNAEAEIKEGISSGIFGLILYLNDNLEECYKLKINLLGLDSLAHVNGVEIGI
NLPIEYAVSMEDKIIQETGRKETRNKIVVDISSLNGGAPGRTRGILVDAIKF"
     repeat_region   7600..9670
                     /gene="At2g41870"
     mRNA            join(7600..8640,9217..9670)
                     /gene="At2g41870"
     CDS             complement(7944..7980)
                     /rpt_family="(GAA)n"
                     join(8017..8640,9217..9417)
                     /gene="At2g41870"
                     /codon_start=1
                     /product="putative DNA binding protein"
                     /protein_id="AAM14826.1"
                     /db_xref="GI:20196896"
                     /translation="MLIYHQERSPDATSNDRDETPETVVREVHALTPAPEDNSRTMT
                     ATLPPPAFRGYFSPPRSAATTMSEGENTTISREFNALVIACSSMENNELMTRDVTQR
                     EDEROEDELMRIHEDTDHEEETNPLAIVPDQYPGSGLDPGCSDNGPGCQSRVGSTVQRVKR
                     EEVEAKITPAWQTAKLAKINNRRKREDAVIINGFNEQVNKANSMMKKIERKLEERKAKA
                     MEKTQNNVAKAQRKAEEBRATAEAAKRGTEVAKVVEVANLMRALGRPPAKRSFSFS"
     repeat_region   complement(8642..8664)
                     /rpt_family="At_rich"
     repeat_region   complement(10329..10357)
                     /rpt_family="At_rich"
     repeat_region   10750..12890
                     /gene="At2g41860"
                     /note="T11A7.4"
     mRNA            join(<10750..11250,11362..11485,11563..11706,11798..11950,
                     12044..12327,12415..12645,12735..>12890)
                     /gene="At2g41860"
     CDS             join(10750..11250,11362..11485,11563..11706,11798..11950,
                     12044..12327,12415..12645,12735..12890)
                     /gene="At2g41860"
                     /codon_start=1
                     /product="putative calcium-dependent protein kinase"
                     /protein_id="AAM14824.1"
                     /db_xref="GI:20196892"
                     /translation="MGNCCGTAGSLIQDKOKKGKPLNPFSNEYGNHHDGLKLIVLKE
                     PTGHEIKOKVKLGRELGRGEFGVTYLCTETETGELFACKSILKKLKTSIDIEDVKRE
                     VEIMRQMEHPNIVTLKETYEDRAVHLVWELCEGGELFDRIVARGHYTERAAASVIR
                     TIIEVVQMCHKHGVMHRDLKPENFLFANKKETSLKAIDFGLSVFFKPGERFNEIVGS
                     PYMAPEVLRRSYGQEIDIWSAGVLIYILLCGVPPFWAETEHGVAKAILKSVIDFKRD
                     PWPKVSDNAKDLIKKMLHPDPRRLTAQOVLDHPWIQNGKMAPNSNVSLGETVRARLKOF
                     SVMNKLKRRALRVIAEHLSVEETSCIKEROVMDTSNRGKITTTELGIGLQKLGIVVP
                     QDDIQILMDAGDVDKDGTLDVNEFVAISVHIRKLGNDEHLKKAFTFFDKNKSGYIEIE
                     LRDALADVDVDTTSEEVVEAIILDVDTNKDGKISYDEFATMMKRGTDWRKASROYSRD
                     LFKCLSLKLMQDGSLQSNGDTK"
     repeat_region   complement(14214..14243)
                     /rpt_family="At_rich"
     repeat_region   complement(14694..14717)
                     /rpt_family="At_rich"
     repeat_region   complement(15877..15901)
                     /rpt_family="At_rich"
     gene            16477..18618
                     /gene="At2g41850"
                     /note="T11A7.5"
     mRNA            join(<16477..16731,16830..16961,17046..17216,17458..17665,
                     17823..17904,17999..18107,18206..18319,18409..>18618)
                     /gene="At2g41850"
     CDS             join(16477..16731,16830..16961,17046..17216,17458..17665,
                     17823..17904,17999..18107,18206..18319,18409..18618)
                     /gene="At2g41850"
                     /codon_start=1
                     /product="putative polygalacturonase"
                     /protein_id="AAC02763.1"
                     /db_xref="GI:2335094"
                     /translation="MARCTNLVTVFLLWALLMFSVCKASRISPNVYDHSYKRFKSDSL
                     IKRREDITGLRSFVRASLRFPTTVSVSDFGAKGDGKTDDYQAFVNAWKRACSSNGAVN
```

```
RESULT 11
LOCUS       AC104652          130206 bp    DNA     linear   PRI 21-FEB-2002
DEFINITION  Homo sapiens BAC clone RP11-273F23 from 4, complete sequence.
ACCESSION   AC104652
VERSION     AC104652.3  GI:18464293
KEYWORDS    HTG.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 130206)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 130206)
  AUTHORS   Nguyen,C., Haakenson,W. and Boyer,E.
  TITLE     The sequence of Homo sapiens BAC clone RP11-273F23
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 130206)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-DEC-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 130206)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 130206)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Feb 1, 2002 this sequence version replaced gi:18087701.
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            ----------  Summary Statistics
            Center project name: H_NH0273F23
            Drafting Center: WIBR
            --------------------------
            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.
            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.
            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc
            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR: pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is RP11-679C8, 2000 bp overlap.
            The actual end of this clone is at base position 1 of RP11-679C8.
            Actual start of this clone is at base position 61045 of RP11-273F23;
            actual end is at base position 61045 of RP11-273F23.
```

```
Qy       41604  AGTTAAACGTAAAGCTTTATTAAACAAAAGAACAAAT  41639
                |||| ||     |||||||| |||||||  |||| ||
Db         226  TGTTGAAAAGAGACCTTTGGCACACATTGAAGGGAT  261

Qy       41544  CATTTAATTATAAATCTTATTGACACAAATTTCTTTGGTTTAACGTTTTTTTTTCTTTGTA  41603
                ||                 | | || |  |||| |  | |||||
Db         166  CTTTTGGAATAAAAAACCTCCTCGGGAACTTCCGCTTAACATATACTCTCCGTGTACACATGCG  225

Qy       41484  GGAGACCAATATATGGACTTCCGGGAACCCCGTGTTCTTTCCGTTGTTTAATCTAAATGTT  165
                || ||    |      | |  ||| |||||  |||||||| ||| ||| |||||   | ||
Db         106  GGAGCTGAAGGATGCAGCGGTTGTTAAGCCCCGTGTTCTTTCCGTTGTTTAATCTAAATGTT  165

Query Match     10.4%;  Score 39.2;  DB 8;  Length 91071;
Best Local Similarity  53.2%;
Matches   83;  Conservative   0;  Mismatches   73;  Indels   0;  Gaps   0;
                        Pred. No. 0.76;
```

The sequence of AC034273 has been incorporated into AC104652.

```
FEATURES             Location/Qualifiers
     source          1..130206
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-273F23"
                     /clone_lib="RPCI-11"
     repeat_region   1341..1625
                     /rpt_family="Alu"
     repeat_region   1623..1677
                     /rpt_family="(TAAA)n"
     repeat_region   5763..5808
                     /rpt_family="L2"
     repeat_region   5950..5979
                     /rpt_family="AT_rich"
     repeat_region   6514..6545
                     /rpt_family="AT_rich"
     repeat_region   6564..6638
                     /rpt_family="AT_rich"
     repeat_region   7741..7919
                     /rpt_family="L2"
     repeat_region   8052..8084
                     /rpt_family="MIR"
     repeat_region   8194..8491
                     /rpt_family="L2"
     repeat_region   8625..8992
                     /rpt_family="MaLR"
     repeat_region   9173..9245
                     /rpt_family="L2"
     repeat_region   9246..9985
                     /rpt_family="MaLR"
     repeat_region   9962..9996
                     /rpt_family="L1"
     repeat_region   9986..10388
                     /rpt_family="AT_rich"
     repeat_region   11434..11649
                     /rpt_family="MIR"
     repeat_region   11841..11877
                     /rpt_family="AT_rich"
     repeat_region   12133..12461
                     /rpt_family="MER1_type"
     repeat_region   13822..14118
                     /rpt_family="Alu"
     repeat_region   14950..15212
                     /rpt_family="Alu"
     repeat_region   17097..17679
                     /rpt_family="ERVL"
     repeat_region   17680..17719
                     /rpt_family="ERVL"
     repeat_region   18172..22043
                     /rpt_family="L1"
     repeat_region   22043..22081
                     /rpt_family="(T)n"
     repeat_region   22250..22278
                     /rpt_family="AT_rich"
     repeat_region   22263..22476
                     /rpt_family="L1"
     repeat_region   22477..22755
                     /rpt_family="L1"
     repeat_region   22829..23213
                     /rpt_family="L1"
     repeat_region   23269..23460
                     /rpt_family="L1"
     repeat_region   23462..23508
                     /rpt_family="(TAGA)n"
     repeat_region   23577..23811
                     /rpt_family="ERVl"
     repeat_region   23820..24408
                     /rpt_family="ERVl"
     repeat_region   24408..25160
                     /rpt_family="ERVl"
     repeat_region   25161..25350
                     /rpt_family="ERV1"
     repeat_region   25351..25487
                     /rpt_family="ERVL"
     repeat_region   25580..26611
                     /rpt_family="ERVL"
     repeat_region   25731..25753
                     /rpt_family="L1"
     repeat_region   26659..27742
                     /rpt_family="AT_rich"
     repeat_region   28051..28355
                     /rpt_family="L2"
     repeat_region   28334..28355
                     /rpt_family="Alu"
     repeat_region   28435..29100
                     /rpt_family="(A)n"
     repeat_region   29101..29571
                     /rpt_family="L2"
     repeat_region   29621..29797
                     /rpt_family="L2"
     repeat_region   29713..29742
                     /rpt_family="L2"
     repeat_region   29801..29873
                     /rpt_family="AT_rich"
     repeat_region   30324..30360
                     /rpt_family="L2"
     repeat_region   30687..30735
                     /rpt_family="(TG)n"
     repeat_region   30736..30767
                     /rpt_family="(TC)n"
     repeat_region   30958..31356
                     /rpt_family="(CA)n"
     repeat_region   31571..31598
                     /rpt_family="ERV1"
     repeat_region   32021..32199
                     /rpt_family="AT_rich"
     repeat_region   32880..32980
                     /rpt_family="MER1_type"
     repeat_region   33097..34152
                     /rpt_family="CT-rich"
     repeat_region   34315..34356
                     /rpt_family="AT_rich"
     repeat_region   34982..35058
                     /rpt_family="AT_rich"
     repeat_region   35084..35119
                     /rpt_family="CR1"
                     /rpt_family="AT_rich"
```

```
RESULT 12
AC108866
LOCUS       AC108866    152822 bp    DNA    linear    PRI 20-APR-2002
DEFINITION  Homo sapiens BAC clone RP11-44D21 from 4, complete sequence.
ACCESSION   AC108866
VERSION     AC108866.5  GI:19848440
KEYWORDS    HTG.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 152822)
  AUTHORS   Sulston,J.E. and Waterston,R.
```

```
Query Match             10.2%;  Score 38.4;  DB 9;  Length 130206;
Best Local Similarity   58.9%;
Matches   66;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy         92 TGAATGTGTGTTCTGGAGCGTGAAAGGATGCACGGTTGTTAAGCCCCTGTTCTTTTCCGTTG 151
              |||||| |  || ||| |||   || |||  | |  ||||||| |||||||||||| | ||
Db     112310 TGAATGATTATATTCTTAGTTAGGTAGGAGAGCCTGGTTTAGTATCCCCTGTTTTTTTCTGTAA 112369

Qy        152 TTTAATCTAAATGTTCTTTGGAATAAAAAACCTCCCGCCAAGTAGTTACTTGGT 203
              ||| ||||| |  |||| |||||||  || |||||| | |||||| || || |
Db     112370 CTTACTCTAAAATGGTTTTGGAAAAATAAACATGTCTAATTAATTCATAATTTAGT 112421
```

```
TITLE     Toward a complete human genome sequence
JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE   99063792
PUBMED    9847074
REFERENCE 2  (bases 1 to 152822)
AUTHORS   Abbott,S., Cotton,M. and Spalding,L.
TITLE     The sequence of Homo sapiens BAC clone RP11-44D21
JOURNAL   Unpublished (2001)
REFERENCE 3  (bases 1 to 152822)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (31-JAN-2002) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE 4  (bases 1 to 152822)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (21-MAR-2002) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE 5  (bases 1 to 152822)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (30-MAR-2002) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE 6  (bases 1 to 152822)
AUTHORS   Waterston,R.H
TITLE     Direct Submission
JOURNAL   Submitted (20-APR-2002) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT   On Mar 30, 2002 this sequence version replaced gi:19570183.
          -------------- Genome Center
          Center: Washington University Genome Sequencing Center
          Center code: WUGSC
          Web site: http://genome.wustl.edu/gsc
          Contact: saplens@watson.wustl.edu
          -------------- Summary Statistics
          Center project name: H_NH0044D21
          --------------

          NOTICE:  This sequence may not represent the entire insert of this
          clone.  It may be shorter because we only sequence overlapping
          clone sections once, or longer because we provide a small overlap
          between neighboring data submissions.

          This sequence was finished as follows unless otherwise noted:
          all regions were double stranded, sequenced with an alternate
          chemistry, or covered by high quality data (i.e., phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by sequence
          from more than one subclone; and the assembly was confirmed by
          restriction digest.

          MAPPING INFORMATION:
          Mapping information for this clone was provided by Dr. John D.
          McPherson, Department of Genetics, Washington University, St. Louis
          MO. For additional information about the map position of this
          sequence, see http://genome.wustl.edu/gsc

          SOURCE INFORMATION:
          The RPCI-11 human BAC library was made from the blood of one male
          donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
          Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
          approach for construction of bacterial artificial chromosome
          libraries. Genomics 51:1-8. The clone may be obtained either from
          Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
          and coworkers at http://www.chori.org
          VECTOR: pBACe3.6

          NEIGHBORING SEQUENCE INFORMATION:
          Actual start of this clone is at base position 1 of RP11-44D21;
          actual end is at base position 152822 of RP11-44D21.
```

```
FEATURES             Location/Qualifiers
     source          1..152822
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-44D21"
                     /clone_lib="RPCI-11"
     repeat_region   6..98
                     /rpt_family="MaLR"
     repeat_region   243..393
                     /rpt_family="MIR"
     repeat_region   1632..1748
                     /rpt_family="L1"
     repeat_region   1747..1819
                     /rpt_family="L1"
     repeat_region   1933..2538
                     /rpt_family="L1"
     repeat_region   2544..2608
                     /rpt_family="ERV1"
     repeat_region   2633..2993
                     /rpt_family="L1"
     repeat_region   3286..3445
                     /rpt_family="MIR"
     repeat_region   5195..5225
                     /rpt_family="(T)n"
     repeat_region   5315..5396
                     /rpt_family="L2"
     repeat_region   6758..7060
                     /rpt_family="Alu"
     repeat_region   8193..8349
                     /rpt_family="MER1_type"
     repeat_region   8366..8496
                     /rpt_family="L2"
     repeat_region   936..9454
                     /rpt_family="MIR"
     repeat_region   10509..10532
                     /rpt_family="AT_rich"
     repeat_region   15923..16255
                     /rpt_family="Alu"
     repeat_region   20537..20557
                     /rpt_family="AT_rich"
     repeat_region   21092..21123
                     /rpt_family="MaLR"
     repeat_region   22318..22496
                     /rpt_family="MIR"
     repeat_region   22629..22661
                     /rpt_family="L2"
     repeat_region   22771..22936
                     /rpt_family="MaLR"
     repeat_region   22991..23068
                     /rpt_family="MaLR"
     repeat_region   23202..23569
                     /rpt_family="L2"
     repeat_region   23750..23822
                     /rpt_family="MaLR"
     repeat_region   23823..24562
                     /rpt_family="L1"
     repeat_region   24563..24965
                     /rpt_family="MaLR"
     repeat_region   26011..26226
                     /rpt_family="MIR"
     repeat_region   26418..26454
                     /rpt_family="AT_rich"
     repeat_region   26710..27038
                     /rpt_family="MER1_type"
     repeat_region   28399..28695
                     /rpt_family="Alu"
     repeat_region   29527..29789
                     /rpt_family="Alu"
     repeat_region   30235..30280
                     /rpt_family="CR1"
     repeat_region   31674..32256
```

```
RESULT 13
LOCUS       AC080170/c               190912 bp    DNA    linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 10 clone RP11-435B15 map 10, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
ACCESSION   AC080170
VERSION     AC080170.2  GI:12830207
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 190912)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 10, clone RP11-435B15
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 190912)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heeford,A., Horton,L.,
            Illev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,R., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            Meldrim,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neill,D., Oliver,J.,T.M. Oliver,J., Peterson,T.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigillo,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zalnoun,J.,
            Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (28-SEP-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 190912)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heeford,A., Horton,L., Hulme,W., Illev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Trigillo,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,M., Wu,X., Wyman,D., Ye,W.J., Young,G., Zalnoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 14, 2001 this sequence version replaced gi:10334889.
            All repeats were identified using RepeatMasker:
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------------- Genome Center
            Center: Whitehead Institute/MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -------------------- Project Information
            Center project name: L10588
            Center clone name: 435_B_15
            -------------------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 184508 bases at least Q40
            Consensus quality: 187307 bases at least Q30
            Consensus quality: 188400 bases at least Q20
            Insert size: 198000; agarose-fp
            Insert size: 189212; sum-of-contigs
            Quality coverage: 4.9 in Q20 bases; agarose-fp
            Quality coverage: 5.1 in Q20 bases; sum-of-contigs
            --------------------

Query Match   10.2%;   Score 38.4;   DB 9;   Length 152822;
Best Local Similarity   58.9%;   Pred. No. 1.3;
Matches  66;   Conservative  0;   Mismatches  46;   Indels  0;   Gaps  0;

Qy       92  TGAAGTGTGTGTTCTGGAGCTGAAGGAGTGCCACGGTTGGTTGTTAAGCCCCTGTTCTTTTCCGTTG  151
             |||||| || || ||||||||||||| | ||| |||||||||||||| |||||||| |||||| |||
Db   120753  TGAATGATTATATTCTAGTTAGGTAGGTAGAGACCCTGGTTTTAGTATCCCCTGTTTTTTCTGTAA  120812

Qy      152  TTTTAATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTACTTGGT  203
             ||||| ||||| | ||||||| | ||||||| |||| ||||| ||| || ||||
Db   120813  CTTACTCTAAAATGGTTTGGAAAAATAACATGTCTAATAATTCATATATTAGT  120864


repeat_region   32257..32296
                /rpt_family="ERVL"
repeat_region   36621..36659
                /rpt_family="ERVL"
repeat_region   36842..37055
                /rpt_family="(T)n"
repeat_region   37055..37334
                /rpt_family="L1"
repeat_region   37408..37792
                /rpt_family="L1"
repeat_region   37872..38039
                /rpt_family="L1"
repeat_region   38041..38088
                /rpt_family="L1"
repeat_region   38156..38390
                /rpt_family="(TAGA)n"
repeat_region   38399..38987
                /rpt_family="ERVL"
repeat_region   38987..39739
                /rpt_family="ERV1"
repeat_region   39740..39929
                /rpt_family="ERV1"
repeat_region   39930..40066
                /rpt_family="ERVL"
repeat_region   40159..41190
                /rpt_family="L1"
repeat_region   41238..42323
                /rpt_family="L2"
repeat_region   42632..42929
                /rpt_family="L2"
repeat_region   43009..43674
                /rpt_family="Alu"
repeat_region   43675..44371
                /rpt_family="L2"
repeat_region   44375..44447
                /rpt_family="L2"
repeat_region   44898..44934
                /rpt_family="(TG)n"
repeat_region   45261..45307
                /rpt_family="(TC)n"
repeat_region   45308..45339
                /rpt_family="(CA)n"
```

```
*  NOTE: This is a 'working draft' sequence. It currently
*  consists of 18 contigs. The true order of the pieces
*  is not known and their order in this sequence record is
*  arbitrary. Gaps between the contigs are represented as
*  runs of N, but the exact sizes of the gaps are unknown.
*  This record will be updated with the finished sequence
*  as soon as it is available and the accession number will
*  be preserved.
*
*        1     8804: contig of 8804 bp in length
*     8805     8904: gap of 100 bp
*     8905     9991: contig of 1087 bp in length
*     9992    10091: gap of 100 bp
*    10092    11974: contig of 1883 bp in length
*    11975    12074: gap of 100 bp
*    12075    14089: contig of 2015 bp in length
*    14090    14189: gap of 100 bp
*    14190    15697: contig of 1508 bp in length
*    15698    15797: gap of 100 bp
*    15798    18861: contig of 3064 bp in length
*    18862    18961: gap of 100 bp
*    18962    21671: contig of 2710 bp in length
*    21672    21771: gap of 100 bp
*    21772    23839: contig of 2068 bp in length
*    23840    23939: gap of 100 bp
*    23940    28783: contig of 4844 bp in length
*    28784    28883: gap of 100 bp
*    28884    33590: contig of 4707 bp in length
*    33591    33690: gap of 100 bp
*    33690    42274: contig of 8584 bp in length
*    42275    42374: gap of 100 bp
*    42375    51118: contig of 8744 bp in length
*    51119    51218: gap of 100 bp
*    51218    63841: contig of 12623 bp in length
*    63842    63941: gap of 100 bp
*    63942    76817: contig of 12876 bp in length
*    76818    76917: gap of 100 bp
*    76918    90127: contig of 13210 bp in length
*    90128    90227: gap of 100 bp
*    90228   136743: contig of 46516 bp in length
*   136744   136843: gap of 100 bp
*   136844   173143: contig of 36300 bp in length
*   173144   173243: gap of 100 bp
*   173244   190912: contig of 17669 bp in length.

FEATURES             Location/Qualifiers
source               1..190912
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10"
                     /clone="RP11-435B15"
                     /clone_lib="RPCI-11 Human Male BAC"
misc_feature         1..8804
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
misc_feature         8905..9991
                     /note="assembly_fragment"
misc_feature         10092..11974
                     /note="assembly_fragment"
misc_feature         12075..14089
                     /note="assembly_fragment"
misc_feature         14190..15697
                     /note="assembly_fragment"
misc_feature         15798..18861
                     /note="assembly_fragment"
misc_feature         18962..21671
                     /note="assembly_fragment"
misc_feature         21772..23839
                     /note="assembly_fragment"
misc_feature         23940..28783
                     /note="assembly_fragment"
misc_feature         28884..33590
                     /note="assembly_fragment"
misc_feature         33691..42274
                     /note="assembly_fragment"
misc_feature         42375..51118
                     /note="assembly_fragment"
misc_feature         51219..63841
                     /note="assembly_fragment"
misc_feature         63942..76817
                     /note="assembly_fragment"
misc_feature         76918..90127
                     /note="assembly_fragment"
misc_feature         90228..136743
                     /note="assembly_fragment"
misc_feature         136844..173143
                     /note="assembly_fragment"
misc_feature         173244..190912
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    57903 a   33982 c   34268 g   63050 t   1709 others
ORIGIN
```

```
Query Match           10.2%;   Score 38.4;  DB 2;  Length 190912;
Best Local Similarity  58.9%;   Pred. No. 1.3;
Matches   66;  Conservative   0;  Mismatches   46;  Indels   0;  Gaps   0;

Qy     92 TGAATGTGTTGTTCTGGAGCTGAAGGATCGCACGGTTGTTGTPAAGCCCCTGGTTCTTTCCGGTTG 151
          |||||||  | |||| ||||   ||||||||| ||||||||| |||||||||| ||||||||| ||
Db 136670 TGAATGATTATAATTCTAGTTAGGTAGGTGAGAGGCCTGGTTTTAGTATCCCTGGTTTTTCTGTAA 136611

Qy    152 TTTAATCTAATGTTGTTCTTTTGGAATAAAAAACCTCCCTGCCCAAGTACTACTTGGT 203
          ||||| |||  ||| ||||||| | ||||||| | || |||| || || | |||| |||
Db 136610 CTTACTCTAAAATGGTTTGGAAAAATAACATGTCTAATAATTCATTATATTAGT 136559
```

```
RESULT 14
AC024785/c
LOCUS       AC024785            30960 bp    DNA     linear   INV 24-MAY-2002
DEFINITION  Caenorhabditis elegans cosmid Y46C8AL, complete sequence.
ACCESSION   AC024785
VERSION     AC024785.1  GI:7140340
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 30960)
  AUTHORS   Waterston,R.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
   PUBMED   9851916
REFERENCE   2  (bases 1 to 30960)
  AUTHORS   Cordes,M.
  TITLE     The sequence of C. elegans cosmid Y46C8AL
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 30960)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 30960)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 30960)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
```

```
REFERENCE    6 (bases 1 to 30960)
  AUTHORS    Waterston,R.
  TITLE      Direct Submission
  JOURNAL    Submitted (24-MAY-2002) Department of Genetics, Washington
             University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
             Louis, MO 63110, USA
Submitted by:
             Genome Sequencing Center
             Department of Genetics, Washington University
             St. Louis, MO 63110, USA; and
             Sanger Centre, Hinxton Hall
             Cambridge CB10 1RQ, England
             email: rw@nematode.wustl.edu and jes@sanger.ac.uk

COMMENT      NOTICE:  This sequence may not be the entire insert of this clone.
             It may be shorter because we only sequence overlapping sections
             once, or longer because we provide a small overlap between
             neighboring submissions.

             This sequence was finished as follows unless otherwise noted: all
             regions were double stranded, sequenced with an alternate chemistry
             or covered by high quality data (i.e., phred quality >= 30); an
             attempt was made to resolve all sequencing problems, such as
             compressions and repeats; all regions were covered by sequence from
             more than one m13 subclone.

             For a graphical representation of this cosmid sequence and its
             analysis see:
             http://www.wormbase.org/db/seq/sequence?name=Y46C8AL;class=Sequence

                       NEIGHBORING COSMID INFORMATION

             The 5' cosmid is F56D6, 200 bp overlap; the 3' cosmid is Y46C8AM,
             200 bp overlap.

NOTES:

             Coding sequences below are the result of integration and manual
             review of the following data : computer analysis using the program
             Genefinder (P. Green and L. Hillier, personal communication), the
             large scale EST projects of Yuji Kohara
             (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
             elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/),
             similarity to other proteins from BlastX analyses
             (http://blast.wustl.edu/), sequence conservation with C. briggsae
             using Jim Kent's WABA alignment program (Genome Research
             10:1115-1125, 2000), individual C. elegans GenBank submissions,
             and personal communications with C. elegans researchers. tRNAs
             are predicted using the program tRNAscan-SE (Lowe, T.M. and
             Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES             Location/Qualifiers
     source          1..30960
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="IV"
     gene            179..4984
                     /clone="Y46C8AL"
                     /gene="Y46C8AL.3"
                     /note="for a graphical representation of this gene see:
                     http://www.wormbase.org/db/seq/sequence?name=Y46C8AL.3;cla
                     ss=Sequence"
     CDS             join(1479..1698,2460..2723,2768..2836,2897..2951,
                     3004..3097,4144..4383,4441..4585,4636..4733,4784..4905,
                     4951..4984)
                     /gene="Y46C8AL.3"
                     /codon_start=1
                     /product="Hypothetical protein Y46C8AL.3"
                     /protein_id="AAF60598.1"
                     /db_xref="GI:7331910"
                     /translation="MLLLKIFLSTLAGLASAVCPDSNDHEVQGFCFKFEVAQQMTYTD
                     ARNWCHYKNPVGSSYLAYVPDQKTSNYLAFYARTAFGPAAQHFWIGLSKNGSGSLTW
                     DNGSPVGYTNLGSQNGNNLYFTESLANTKWNTLGDDKINYFVCSYNPATTPTTPSTTT
```

```
                     TADKNCQPGQGQTVLFAYSNDLDPISRFDLRQPEDIGYFDYDNVTDYVDSHAPDSSL
                     GFGDTKTGSNVLDVIDKFYDNTKLTPCGSIYMVLLKRYSNSNDISNIVAKRKHGSV
                     NFIASNTPSGGTNSRVLFDLSSRTNGLYVIDRDSYFMQSIDMPLIERYPIYIANPVV
                     TGAGFKDLPLLEVPSFRKHLIMVAVQDRIPVTNVHHVTVQWENSNSEYSCKLDMIPET
                     WNNPYTNANGKRANIDASGFAMTVEYVYNDSGDHPMQIRFYSPEATDFWLPYTD"
     gene            5249..8670
                     /gene="Y46C8AL.4"
                     /note="for a graphical representation of this gene see:
                     http://www.wormbase.org/db/seq/sequence?name=Y46C8AL.4;cla
                     ss=Sequence"
     CDS             join(5249..5341,6210..6426,6506..6784,6836..7019,
                     7066..7159,7390..7623,8127..8271,8322..8416,8469..8590,
                     8637..8670)
                     /gene="Y46C8AL.4"
                     /codon_start=1
                     /product="Hypothetical protein Y46C8AL.4"
                     /protein_id="AAF60599.1"
                     /db_xref="GI:7331911"
                     /translation="MRAEISYPRSAFASFNFQOHLFSLFSLFQLFSMILKLCLLATLFG
                     ITNAACPDSSDQEIRGFCFRFVVQKMTYTDARNWCHYQNPVGPSYLAVVGNQETSNNL
                     AFYARSAAFGQSAAESFWIGLSRNGSGSSLSWDNGFPVIYTNEGSKYGVNNYYFTEKINSK
                     WVTPGDNETNYFVCSYDPTVQPVTQKATTPCTTTPAANNCQFGAROTVLFAYSNDFE
                     PYIVQNTFSNILYKQQVTFAIIREDTRQPQSIMFFNDYNQALAYVRNHLPDYKLGFQ
                     ESTTGSSDVLDVINRYNNSSACASVVMLSKRYPNTPDISRTDIYNTIDYVDSDYHEVY
                     NVPGGSQSQVLFDLASQTNGLYIGIEQDSLFSKRILYMPLRTRYPIYAVNAKVSEGF
                     QVLPEMSVPQDDFLIMVSVQSHLPVSNVQVVNLKWNPSPPYSDKFEMQPVYWDSNT
                     NYNLTIGLSONVYNMTIDIYITNDVETMQIRLSSYFTNYWLPYSN"
     gene            complement(8898..10334)
                     /gene="Y46C8AL.2"
                     /note="for a graphical representation of this gene see:
                     http://www.wormbase.org/db/seq/sequence?name=Y46C8AL.2;cla
                     ss=Sequence"
     CDS             complement(join(8898..9932,10290..10334))
                     /gene="Y46C8AL.2"
                     /codon_start=1
                     /product="Hypothetical protein Y46C8AL.2"
                     /protein_id="AAF60597.1"
                     /db_xref="GI:7331909"
                     /translation="MYDDAHNWCHFKNPVGNKANARYNEVYTDYSDYHQVYSDYSDY
                     HEVYSNYSDYHEVYSNYSDYHEVYSNYSDYHEVYSDYHEVYSNYSDYHQVYSDYHEVYTN
                     YSDYSDYHEVYSNYSDYHEVYSNYSDYHEVYSNYSDYHEVYSNYSDYHEVYSDYSDYHE
                     VYSNYSDYSDYHEVYSNYSDYHEVYSNYSDYHEVYSNYSDYHEVYSNYSDYHEVYSNYS
                     DYSDYHEVYTNYSDYSDYHEVYTDYSDYHEVYSNYSDYHEVYSNYSDYHEVYSDYSDYHEVY
                     SNYSDYSDYHEVYTDYSDYHEVYTDYSVIHEVFSNYSDYHEDFRYHRLKILSIILISI
                     NELFVCYHENMCIISQVLSECKLKF"
     gene            12224..15951
                     /gene="Y46C8AL.5"
                     /note="for a graphical representation of this gene see:
                     http://www.wormbase.org/db/seq/sequence?name=Y46C8AL.5;cla
                     ss=Sequence"
     CDS             join(12224..12333,12381..12503,12615..12780,13562,13778,
                     13870..14145,14204..14387,14441..14534,15204..15437,
                     15486..15630,15678..15772,15826..15951)
                     /gene="Y46C8AL.5"
                     /codon_start=1
                     /product="Hypothetical protein Y46C8AL.5"
                     /protein_id="AAF60602.1"
                     /db_xref="GI:7331914"
                     /translation="MYPENIILAPIVCFFQGIHPKLSLLQQLQHRFQKSLISKLQYFS
                     LQMEQFREIHGTVTEKLRNWILIISKSRTDPISIIMNRKTITPLLPLIAADLPPSSLF
                     FAHQESQFFPLGHFGGRKVGKRGEVDTKCFRFLKLFTLATLLGIAIAACPDSNDREI
                     RGFCEKFVVQKMTYNDARNWCHYQNPVGPSYLAVVGNKETNNLAVYARSAFGASAEY
                     FWIGLSRNGSSGSLSWDNGPFPVIYNEGSHVGNNYFTEKISNTKWDTPGDNEKNYFVC
                     SYDPVVQPVVPKATPTTTNAANVNCQLGAGOTVLFAYSYLAPSVWLNTFSNSLNFE
                     QLVTIAISRDTRQPQSIMFFSDYIQAYTVYSSHLPDSTLGFGDSTTGSNVLDVINF
                     YNDISSCGSSVVMVLSKRYPNTMDISSTVAKVROYHGMVNFLASNARPSGGTQSRVLFDL
                     SSRTNGIYISIEEDATPLHFIGWMPFRERPFIYAVNVKVSGHGSQVLLPPMSVPRDTYL
                     IWVSLQSHLPISNVQSAKLWNFNPPFPQGALSMQPSEWAYTNSNSGGNRRIIDPAVYN
                     MTIDIYITNDVETMQIRFYSSL"
     gene            16234..21008
                     /gene="Y46C8AL.6"
                     /note="for a graphical representation of this gene see:
                     http://www.wormbase.org/db/seq/sequence?name=Y46C8AL.6;cla
```

```
CDS             ss=Sequence"
                join(16234..16362,17175..17391,17491..17739,18508..19286,
                19338..19628,20472..20616,20666..20763,20809..20930,
                20975..21008)
                /gene="Y46C8AL.6"
                /codon_start=1
                /product="Hypothetical protein Y46C8AL.6"
                /protein_id="AAF60603.1"
                /db_xref="GI:7331915"
                /translation="MRASRPHAGRPRADAARRCWHEMKCERLLFEADTFGVPQLSM
                LIKSLLLTTLIGLVSVCPDRNDLEVKGLCFTFVAQOMYNDARDMCHYKNPLGSSFL
                TYVPDSVTSNNLASHARIAFGPAAKNFWIGLSKNTSSGGSLSWDNGLPVVTNFGKNIG
                KSYFSEOVSNAKWNTLGDNDTNFFVCSIDSAGNPSTTPEFLRQRSSVPVWISCHDFVE
                TLPQHFGYIEHRRESSKIPRNNQLTQHDHRIPHCQLLDRIISLKLNKNLISWIIAFLTN
                RVFQVKVGDSLSSFKKADCGVPQGAVLSPLLFGIFVNEIPNILPPAIKCKQFADDLKL
                YTAIPSNSNNVHLQKAIETIVEWSKATKLALNNDKTVCISLGRNTEFQYTIENSLI
                TRSTNVRDLGFQIEPNLSFSMHWKKSVMKSKFMIHQIFSNYCSKNTRLHTLLYKIFVR
```

```
Query Match       10.1%;  Score 38.2;  DB 3;  Length 30960;
Best Local Similarity   63.7%;  Pred. No. 1.5;
Matches  58;  Conservative  0;  Mismatches  33;  Indels  0;  Gaps  0;

Qy  155  AATCTAATGTTCTTTGGAATAAAAACCTCCCTGCCAAGTAGTACTTGGTTTTATGCTCAA  214
           ||| ||||||||| || ||| |||||||||||||||| |||||||| | || ||| ||||
Db  1728 AATCCACTGTACGTTAAAATATATTACTTACATGCCAAGTAGTTGCTTGTTTTTTGATCAG  1669

Qy  215  CATGCTTTTGACTGTTGTTGAAAAAGAGACCTTTGG  245
           ||  ||| || || |||||    |||    ||| |||
Db  1668 GAACATATGCCAGATAGGAGAGAGCCTACTGG  1638
```

```
BASE COUNT    59260 a   30996 c   29354 g   54386 t    200 others
ORIGIN
```

```
FEATURES          Location/Qualifiers
source            1..174196
                  /organism="Danio rerio"
                  /db_xref="taxon:7955"
                  /clone="CH211-144F18"
                  /clone_lib="CHORI-211"
     1   53761: contig of 53761 bp in length
         53762  53861: gap of  100 bp
         53862  77324: contig of 23463 bp in length
         77325  77424: gap of  100 bp in length
         77425  174196: contig of 96772 bp in length
misc_feature      1..53761
                  /note="assembly_fragment:00869
                  fragment_chain:1"
misc_feature      53862..77324
                  /note="assembly_fragment:01503
                  fragment_chain:1"
misc_feature      77425..174196
                  /note="assembly_fragment:00262
                  fragment_chain:1"
```

```
RESULT 15
AL732628/c
LOCUS        AL732628     174196 bp    DNA      linear    HTG 09-AUG-2002
DEFINITION   Danio rerio clone CH211-144F18, *** SEQUENCING IN PROGRESS ***, 3
             unordered pieces.
ACCESSION    AL732628
VERSION      AL732628.8   GI:22204545
KEYWORDS     HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE       zebrafish.
ORGANISM     Danio rerio
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
             Cypriniformes; Cyprinidae; Danio.
REFERENCE    1  (bases 1 to 174196)
AUTHORS      Corby,N.
TITLE        Direct Submission
JOURNAL      Submitted (31-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
             Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk
             Clone requests: clonerequest@sanger.ac.uk
             On Aug.11, 2002 this sequence version replaced gi:21911655.
COMMENT      -------------- Genome Center
             Center: Wellcome Trust Sanger Institute
             Center code: SC
             Web site: http://www.sanger.ac.uk
             Contact: zface@sanger.ac.uk
             -------------- Project Information
             Center project name: zC14F18
             -------------- Summary Statistics
             Assembly program: XGAP4; version 4.5
             Chemistry: Dye-terminator; 100% of reads
             Consensus quality: 173595 bases at least Q40
             Consensus quality: 173741 bases at least Q30
             Consensus quality: 173854 bases at least Q20
             Insert size: 173996; sum-of-contigs
             Insert size: 161892; 19.9% error; agarose-fp
             Quality coverage: 6.72x in Q20 bases; sum-of-contigs Quality
             coverage: 7.22x in Q20 bases; agarose-fp
             --------------
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 3 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             *
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
```

```
Query Match       10.0%;  Score 37.8;  DB 2;  Length 174196;
Best Local Similarity   54.7%;  Pred. No. 2;
Matches  75;  Conservative  0;  Mismatches  62;  Indels  0;  Gaps  0;

Qy  200  TGGTTTTATGCTCAACACATGCTTTTGACTGTTGAAAAAGAGACCTTTGGCACACATTGAAGG  259
           ||  ||||| || |||| || | ||||||| |||  ||||| || |||  |||  ||||||||
Db  98366 TGGACGTGTTGTTCAACATGCAATAAAATGTGGATAAGAACCAAGGAAGCACTTTTGAAGGC  98307

Qy  260  ATGGTGAATGGAGAATGCCAATCCATGGAATCAGGTGGGCGGCAGCGTATGTTGGTAGCTATAGC  319
           ||| |||| | ||  |||||||  || |||||| ||  || ||  |||| || || |||| | |
Db  98306 AAGTTCCAGTATAAAACAATTAGTCGCACATCCCCAGGCGCTCCAAAGTTTCATCGCAAATTC  98247

Qy  320  AGAAGTCTTCTTGGCAA  336
           | || ||| || |||
Db  98246 GGGTGTTTTATTTTTAA  98230
```

```
Search completed: January 31, 2003, 03:21:40
Job time : 3850 secs
```